



Estudio de la generación de las fracciones de residuos urbanos por parte de las actividades económicas en la ciudad de Madrid

Entregable 2
*Definición de la población objetivo y diseño de la
muestra*

INFORME BORRADOR, SUJETO A REVISIÓN

Mayo 2024

Un trabajo realizado por:



Afi

Analistas Financieros
Internacionales

ksnet



Índice

1. Introducción: Motivación y objeto del informe	4
2. Estudio de los determinantes de la generación de residuos en la ciudad de Madrid	6
2.1. <i>La variable dependiente: la generación de residuos en la ciudad de Madrid por distritos</i>	7
2.1.1. Construcción de la variable generación de residuos	7
2.1.2. Dato administrativo vs dato estadístico	10
2.1.3. La generación de residuos: hogares y actividades económicas.	12
2.1.4. Análisis descriptivo de la variable dependiente generación de residuos	12
2.2. <i>Las variables independientes del modelo: la realidad sociodemográfica y económica de los distritos</i>	15
2.2.1. La estructura demográfica y económica de los distritos	17
2.2.2. Actividades económicas: registro de locales	22
2.2.3. La influencia del sector turístico en la generación de residuos	27
2.3. <i>Determinantes de la generación de residuos de la Ciudad de Madrid por distrito</i>	29
2.3.1. Modelo de datos de corte transversal 2023	30
2.3.2. Modelo de datos panel 2018- 2023	32
3. La identificación de zonas homogéneas en Madrid: un análisis clúster para la estratificación de los distritos en la ciudad	38
3.1. <i>Descripción metodológica de la técnica de análisis clúster</i>	39
3.2. <i>Análisis secuencial para la identificación de zonas homogéneas</i>	43
3.2.1. Etapa 1. Zonas homogéneas por generación de residuos	44
3.2.2. Etapa 2. Zonas homogéneas por generación de residuos y actividades económicas y del turismo	46
3.2.3. Etapa 3. Zonas homogéneas con la inclusión de todo el pool de variables	48
4. Muestreo aleatorio estratificado por zonas homogéneas	51
4.1. <i>Análisis descriptivo de los ficheros de trabajo iniciales.</i>	51
4.1.1. Análisis descriptivo de Listado_AE y proceso de depuración	51
4.1.2. Análisis descriptivo de MARCO_TRUA y proceso de depuración.	55
4.1.3. Análisis descriptivo de los listados de contenedores	57
4.1.4. Creación del marco muestral: enlace de ficheros.	59
4.2. <i>Proceso de definición de la muestra</i>	60

4.2.1 Alternativa 1. Distribución proporcional de la muestra	62
4.2.2. Alternativa 2. Distribución homogénea de la muestra	64
4.2.3. Alternativa 3. Distribución estableciendo un mínimo	65
4.2.4. Conclusiones sobre la distribución de la muestra.	67
4.3. <i>Proceso de extracción de la muestra</i>	68
4.4. <i>Nueva distribución tras la eliminación de los Grandes Productores.</i>	70

1. Introducción: Motivación y objeto del informe

El presente documento constituye el entregable 2 del proyecto “Estudio de la generación de las fracciones de residuos urbanos por parte de las actividades económicas en la ciudad de Madrid”. El objetivo de este trabajo es **definir cómo se realizará el trabajo de campo** en el que a través de encuestas se determinarán las cantidades en kilogramos que cada tipo de actividad económica genera de cada una de las fracciones de residuos a analizar en este estudio. La definición del trabajo de campo a su vez cuenta con tres objetivos, que son:

- La **identificación de la población objetivo** que constituye el universo a analizar dentro de la población marco (totalidad de negocios de la ciudad de Madrid)
- La **definición estadística de la muestra de establecimientos a entrevistar garantizando la representatividad que se requiere** para la obtención de buenos resultados durante el trabajo de campo.
- Definir la manera y la fuente de la que se extraerá la muestra, dadas las limitaciones detectadas en las bases de datos del censo de locales y la propia Tasa de Residuos Urbanos.

Para alcanzar los objetivos descritos, se ha desarrollado un proceso en tres etapas:

- **Etapla 1. Estudio de los determinantes en la generación de residuos de las actividades económicas**

En esta etapa se identifican las variables que determinan el nivel de generación de residuos y, en base a ellas, los grupos en los que se clasificarán los 21 distritos de la ciudad de Madrid.

La identificación de las variables clave que explican el nivel de generación de residuos se realiza a través de un ejercicio de inferencia estadística a partir de modelos econométricos de regresión.

Posteriormente se construye un conjunto de información de tipo panel, en el que, para cada distrito, se establece el conjunto de información tanto de la variable dependiente (generación de residuos) como de las variables independientes que se han identificado.

- **Etapla 2. Análisis clúster para la identificación de estratos heterogéneos representativos de la población objeto de estudio.**

En el marco del proyecto, se considera que el análisis clúster es una técnica idónea para realizar la identificación de zonas homogéneas dentro de la ciudad de Madrid en la generación de residuos.

Con dicho fin, se ha desarrollado un modelo de clusterización basado en k-medias, que implica la determinación del número de clústeres a priori, pero también jerárquico, para que sea el propio método el que constituya los grupos de zonas homogéneas.

- **Etapas 3. Muestreo aleatorio simple dentro de la estratificación realizada en la Etapa 2 para identificar la población muestral sobre la que realizar el trabajo de campo.**

Dentro de esta etapa se desarrollan dos tareas fundamentales:

- Análisis de diferentes fuentes de información para determinar el número de establecimientos por zona homogénea y actividad. Este análisis permite **establecer la distribución concreta de la población marco**.
- Elaboración de una **distribución de la muestra estratificada** con un número total de establecimientos de 1.200 que forman parte de la muestra final.

El resultado se materializa en una tabla final que recoge, para cada zona homogénea identificada en la etapa clúster, el número de entrevistas que se realizarán en la siguiente fase del trabajo en cada una de las actividades económicas.

2. Estudio de los determinantes de la generación de residuos en la ciudad de Madrid

El objetivo de este apartado es identificar, a partir de un ejercicio de inferencia causal, las variables que más significativamente explican la generación de residuos en la ciudad de Madrid. Esta información se necesita para poder determinar posteriormente si existen zonas homogéneas en Madrid en lo que se refiere a generación de residuos, de forma que puedan estudiarse conjuntamente a través de la encuesta.

El proceso se inicia con la recopilación de los datos sobre el total de generación de residuos y de las variables relacionadas con la estructura demográfica, socioeconómica y vinculadas con las actividades económicas que pueden asociarse con dicha generación. El nivel territorial de referencia adoptado es el correspondiente a los distritos, ya que la información sobre las cantidades de residuos no está disponible por barrio.

Con la información recabada se construye una **base de datos longitudinal**, a nivel de distrito, que abarca los años comprendidos entre 2018 y 2023¹.

Con la base obtenida se desarrolla un tramo de aplicación de **modelos de regresión de datos panel de efectos fijos a nivel individual**. El fin es identificar las variables que mejor definen la generación de residuos, haciendo hincapié en aquellas que pueden afectar a la generación de residuos por parte de las actividades económicas.

A continuación, se expone con detalle el procedimiento seguido en el desarrollo de las tareas presentadas anteriormente, y que abarcan desde la selección y limpieza de las variables, hasta la obtención de los resultados finales.

En ese marco, por un lado, se analizan las características de la variable dependiente: la generación de residuos en la ciudad de Madrid por distritos. Por otro, se estudian las variables que se consideran relacionadas con la generación (variables explicativas). Finalmente, se presentan los resultados que se obtienen con la aplicación de los modelos de regresión, con la correspondiente síntesis de los hallazgos alcanzados.

¹ Si bien la información sobre la generación de residuos se encuentra disponible desde 2017, el panel se construye a partir de 2018 debido a que la mayoría de las variables explicativas no cuentan con datos correspondientes a ejercicios anteriores.

2.1. La variable dependiente: la generación de residuos en la ciudad de Madrid por distritos

El estudio de la generación de residuos por parte de las actividades económicas en la ciudad de Madrid hace preciso conocer las características de dicha variable, a la que se denomina dependiente.

La variable dependiente está determinada por otras que se denominan variables independientes y que, como anteriormente se exponía, se identifican a lo largo del trabajo.

En esta parte del informe se expone el estudio realizado sobre la generación de residuos, descendiendo al detalle por distritos.

2.1.1. Construcción de la variable generación de residuos

El estudio de la variable dependiente requiere, en primer lugar, recopilar los datos correspondientes a la misma. En este sentido, se han identificados **dos fuentes de información**:

1. Una de **carácter estadístico**, de disponibilidad pública, a través del Portal de Datos del Ayuntamiento de Madrid.
2. Otra de **carácter administrativo**, que contiene la información recopilada por el Área de Medioambiente del propio Ayuntamiento. Incluye archivos que recogen datos diversos asociados a la generación: cantidades, tipos y número de cubos, recorridos realizados en la recogida, etc.

La información más valiosa se encuentra, como más adelante se explica, en la fuente de carácter administrativo. En este ámbito, se ha trabajado con diferentes archivos. Los fundamentales están formados por **seis bases de datos**, una por cada año del período analizado (2018-2023), que contienen las cantidades de residuos recogidas.

El **resumen** de la información contenida en cada base de datos proporciona las toneladas recogidas en la ciudad de Madrid dentro de la tipología de servicio que se indica en la siguiente tabla.

Tabla 1. Tipos de servicio-residuos establecidos en la recogida y transporte

Servicio
Resto viviendas y locales
Orgánica viviendas y locales
Envases viviendas y locales
Papel-Cartón viviendas y locales
Vidrios viviendas y locales
Grandes productores resto
Construcción y demolición (RCDs)
Clínicos
Puntos limpios (voluminosos, metales y poda)
Mercados
Grandes productores orgánica
Cartón comercial
Ropa Usada
Aparatos eléctricos y electrónicos (RAEEs)
Vidrio puerta a puerta
Residuos vegetales Migas Calientes
Puntos limpios fijos (excepto muebles madera, metales,RAEEs, poda, papel, vidrio, ropa, RCDs, aceite y pilas)
Madera de puntos limpios fijos
Aceite vegetal
Recogida de animales muertos
Pilas
Puntos limpios de proximidad (excepto RAEEs, pilas, ropa y aceite vegetal)
Saneamiento de viviendas
Puntos limpios móviles (excepto RAEEs, pilas, ropa y aceite vegetal)

Fuente: Ayuntamiento de Madrid

Sin embargo, **el desglose por distritos presenta una clasificación diferente**. Por ejemplo, dentro de la fracción resto, se agrupa la generada por viviendas, locales, mercados², grandes productores clínicos, puntos limpios y animales muertos. En el caso de la fracción orgánica, se incluye, además de viviendas y locales, la asociada a grandes productores; o, respecto a la fracción papel-cartón, destaca que la información por distritos contempla también las cuantías correspondientes a la recogida puerta a puerta.

Para obtener la información más completa posible, también se ha trabajado con tres archivos proporcionados por el Ayuntamiento con datos adicionales sobre la generación de residuos de los **grandes productores y mercados** (Mercamadrid, principalmente) obtenidos de las empresas encargadas de la recogida de residuos en la ciudad (3 lotes territorializados).

Para construir la variable de interés para el estudio – generación de residuos por distrito – se ha cruzado la información entre los datos administrativos de los servicios de recogida del Ayuntamiento de Madrid, y los ficheros adicionales con información de mercados y grandes productores, tal y como se ilustra en el siguiente esquema de trabajo.

² Mercados y Mercamadrid tienen, además, una contabilización específica y global, sin diferenciar fracciones generadas o distrito de ubicación.

SERVICIO	FRACCIÓN	DEFINICIÓN
RECDOMIRES (cubo gris)	RESTO	Son las toneladas de la fracción resto y que se recoge tanto a las viviendas como a los locales. Si entra en el estudio
RECDOMIORG (cubo marrón: carga trasera)	ORGÁNICA	Son las toneladas de la fracción orgánica y que se recoge tanto a las viviendas como a los locales. Si entra en el estudio
RECDOMIENV (cubo amarillo)	ENVASES	Son las toneladas de la fracción envases y que se recoge tanto a las viviendas como a los locales. Si entra en el estudio
PAPEL-CARTÓN (Iglúes y dependencias)	PAPEL-CARTÓN	Son las toneladas de la fracción papel y cartón y que se recoge tanto a las viviendas como a los locales. Si entra en el estudio pero hay que tener en cuenta que estos contenedores no los pueden usar los comercios que tienen la recogida de "Cartón comercial"
RECVIDRIOS (iglú)	VIDRIO	Son las toneladas de la fracción vidrio y que se recoge tanto a las viviendas como a los locales. Si entra en el estudio pero hay que tener en cuenta que estos contenedores no los pueden usar los comercios que tienen la recogida de "Vidrio puerta a puerta"
RECGRANPOD (recogida resto con amplirroll)	RESTO	Son las toneladas de la fracción resto y que se recoge a ciertas empresas. Las tenemos identificadas (ubicación del contenedor y kilos generados por cada una)
RCDs	-	No para el estudio
REC CLÍNICOS (4 ruedas y amplirroll)	-	No para el estudio
RECLIMPIOS (Voluminosos, Metales y Poda)	-	No para el estudio
RECGRANPDORG (orgánica con amplirroll)	ORGÁNICA	Son las toneladas de la fracción orgánica y que se recoge a ciertas empresas. Las tenemos identificadas (ubicación del contenedor y kilos generados por cada una)
Mercados (mercados 4 ruedas y amplirroll)	MERCADOS Y MERCAMADRID	Son las toneladas de la fracción resto y que se recoge a los mercados como apoyo a la recogida normal (la de los primeros cubos). Es decir a estos mercados se les recoge dos veces.
Cartón comercial	PAPEL-CARTÓN	Son las toneladas de la fracción papel y cartón y que se recoge a ciertos locales (tenemos las direcciones de las calles donde tienen que sacar el papel. No hay contenedores)
ROPA USADA	ROPA	Son las toneladas de la fracción ropa usada y que se recoge tanto a las viviendas como a los locales (¿?)
RAAEs	-	No para el estudio
RECVIDRIOPP (vidrio puerta a puerta)	VIDRIO	Son las toneladas de la fracción vidrio y que se recoge a ciertos locales (tenemos las direcciones de las calles donde se ubican los contenedores)
RESIDUOS DEPOSITADAS EN MIGAS CALIENTES	-	No para el estudio
PLF (excepto muebles madera, metales,RAEEs, poda, papel, vidrio, ropa, RCDs, aceite y pilas)	-	No para el estudio
Aceite vegetal	ACEITE	Son las toneladas de la fracción aceite de cocina usado y que se recoge tanto a las viviendas como a los locales (¿?)
MADERA DE PUNTOS LIMPIOS FIJOS	-	No para el estudio
PLP (excepto RAEs, pilas, ropa y aceite vegetal)	-	No para el estudio
Pilas (D.G. Servicios de Limpieza y Residuos)	-	No para el estudio
RECANIMALESM	-	No para el estudio
SANEAMIENTO DE VIVIENDAS	-	No para el estudio
PLM (excepto RAEs, pilas, ropa y aceite vegetal)	-	No para el estudio

Tabla grandes productores

Vs

Tabla resumen

Tipología		RESUMEN RESIDUOS 2023	
		Resumen de datos de la recogida y transporte de residuos urbanos en la ciudad de Madrid	
		Servicio	TOTAL
		RECDOMIRES (cubo gris)	518.021
		RECDOMIORG (cubo marrón: carga trasera)	281.359
		RECDOMIENV (cubo amarillo)	126.390
		PAPEL-CARTÓN (Iglúes y dependencias)	74.146
		RECVIDRIOS (iglú)	60.300
		RECGRANPOD (recogida resto con amplirroll)	29.156
		REC CLÍNICOS (4 ruedas y amplirroll)	16.165
		RCDs	18.771
		RECGRANPDORG (orgánica con amplirroll)	11.872
		Cartón comercial	11.635
		RECLIMPIOS (Voluminosos, Metales y Poda)	11.406
		RAAEs	3.563
		Mercados (mercados 4 ruedas y amplirroll)	3.301
		ROPA USADA	2.718
		RECVIDRIOPP (vidrio puerta a puerta)	2.232
		RESIDUOS DEPOSITADAS EN MIGAS CALIENTES	1.188
		PLF (excepto muebles madera, metales,RAEEs, poda, papel, vidrio, ropa, RCDs, aceite y pilas)	532
		Aceite vegetal	382
		MADERA DE PUNTOS LIMPIOS FIJOS	360
		RECANIMALESM	110
		PLP (excepto RAEs, pilas, ropa y aceite vegetal)	82
		SANEAMIENTO DE VIVIENDAS	74
		Pilas (D.G. Servicios de Limpieza y Residuos)	73
		PLM (excepto RAEs, pilas, ropa y aceite vegetal)	38
			1.143.873
		Color negro: destino PTV	1.030.314
		Color azul: destino no PTV	27.704
		Color rojo: GESTORES DE PAPEL-CARTÓN	88.781
		Color verde: destino PTV pero no con contratos de recogida	74

Como resultado del cruce de información entre las tablas de grandes productores y mercados y la generación de residuos, se obtiene el total agregado de generación de residuos por distritos de la ciudad de Madrid (dato administrativo).

En la mayoría de los distritos, el peso de la generación de residuos por parte de los grandes productores sobre el total oscila entre el 0% y el 5%, con excepción de Villa de Vallecas. No obstante, únicamente se cuenta con esta información para el año 2023.

Las denominaciones de los tipos de servicio-residuo en los tres lotes son diferentes a las que se presentan en las seis bases de datos, anteriormente mencionadas. Por tanto, **ha sido necesario realizar un trabajo de identificación para disponer de un conjunto de información homogéneo y completo.**

Realizadas las consideraciones anteriores, es preciso señalar que el trabajo de análisis de la generación de residuos se ha circunscrito a los siguientes tipos:

- **Resto**
- **Envases**
- **Orgánica**
- **Vidrio**
- **Papel-Cartón³**

La descripción de los datos anteriores permite comprobar que las toneladas de residuos recogidas incluyen tanto las generadas por los hogares/ciudadanía como por las actividades económicas. Por lo tanto, si bien el estudio se centra en las correspondientes a estas últimas, la información existente no ha permitido separarlas.

A partir de los datos recopilados y los ajustes realizados, **se ha construido un panel donde para cada distrito y año se cuenta con los residuos, diferenciados por tipo de servicio económico, generados en toneladas.**

2.1.2. Dato administrativo vs dato estadístico

La información recogida en el Portal de Datos del Ayuntamiento de Madrid es menos completa que la que se obtiene de la fuente administrativa. Esta es la razón por la que se ha trabajado con los datos proporcionados por esta última.

³ El trabajo también contempla las fracciones de ropa y aceite usado, sin embargo, la información de su generación no está disponible a nivel de distrito para el período considerado.

En concreto, la tabla siguiente muestra que, para el año 2023, la proporción que representan los datos estadísticos (Portal de Datos del Ayuntamiento) sobre los administrativos se encuentra por debajo del 100% en todos los distritos, salvo en el de Chamberí.

La situación descrita significa que, en general, las cuantías de residuos consignadas en el Portal de Datos son inferior a las proporcionadas por el Ayuntamiento dentro de la fuente administrativa procedente del Área de Medioambiente.

Tabla 2. Porcentaje de los datos administrativos sobre los datos estadísticos de generación de residuos

Distrito	Proporción 2023 (Estadístico / Administrativo)
Chamberí	100,3%
Carabanchel	99,5%
Villaverde	99,0%
Vicálvaro	98,9%
Puente de Vallecas	98,1%
Salamanca	97,9%
Centro	97,7%
Hortaleza	97,6%
Tetuán	96,5%
Ciudad Lineal	96,1%
Latina	95,9%
Chamartín	95,6%
Moratalaz	95,2%
San Blas-Canillejas	94,8%
Arganzuela	94,6%
Barajas	93,2%
Fuencarral-El Pardo	92,8%
Moncloa-Aravaca	90,9%
Retiro	89,3%
Usera	88,4%
Villa de Vallecas	58,1%

Fuente: Datos del Ayuntamiento de Madrid y del Portal de datos abiertos

El caso más destacado es el de Villa de Vallecas, donde el dato estadístico representa el 58% del administrativo. La causa de este desvío tan significativo puede encontrar su explicación en la actividad asociada a la presencia de Mercamadrid, un gran generador de residuos.

Por tanto, **el trabajo con los datos administrativos permite asumir que los resultados que se alcancen tendrán un mayor grado de validez externa que en el caso de haber desarrollado el análisis con los datos estadísticos.**

2.1.3. La generación de residuos: hogares y actividades económicas.

Al igual que en la generalidad de los municipios españoles, las características del servicio de recogida y transporte de residuos en la ciudad de Madrid determinan que no es posible la diferenciación del volumen de residuos generados por los hogares de aquel correspondiente a las actividades económicas, lo que conforma el objeto de este proyecto.

Esta cuestión es relevante, puesto que puede tener incidencia en los resultados que se obtengan del análisis econométrico de los determinantes de la generación de residuos. Es por ello que se han analizado fuentes que nos permitan contar con una aproximación a esta información. En este sentido, el Banco de Datos del Ayuntamiento de Madrid incluye una estadística sobre la producción anual de residuos urbanos, donde se puede comprobar el total de residuos generados por los hogares y por las actividades económicas dentro de la ciudad de Madrid.

La tabla que se presenta a continuación muestra que las actividades económicas han sido responsables de la generación en entre un 5% y un 8% del total de residuos, en el período comprendido entre 2018 y 2022. Dentro de ese intervalo porcentual, es necesario destacar que el valor se ha ido reduciendo de forma continuada desde el año 2019.

Tabla 3. Porcentaje de los residuos generados por las actividades económicas

Año	% generado por actividades económicas	% generado por ciudadanía
2018	7,98%	92,02%
2019	8,40%	91,60%
2020	6,94%	93,06%
2021	5,40%	94,60%
2022	5,28%	94,72%

Fuente: Datos del Ayuntamiento de Madrid y del Portal de datos abiertos

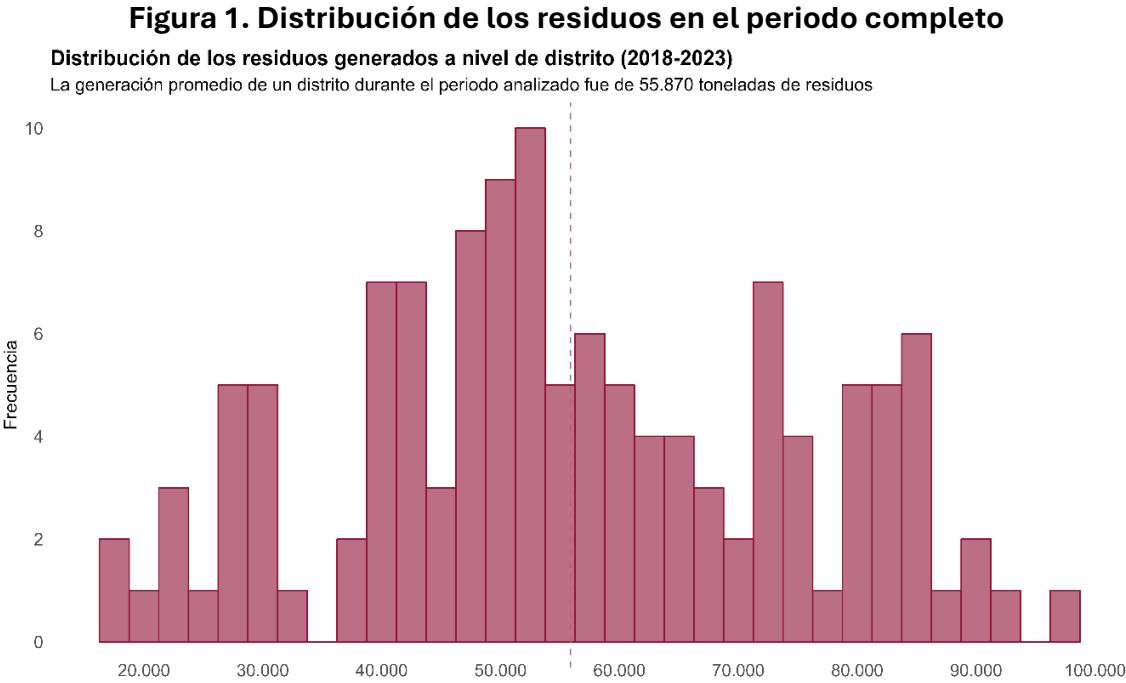
Los aspectos descritos implican que los análisis econométricos y las tareas de clusterización capturarán tanto la variabilidad de los residuos generados por las actividades económicas como por la ciudadanía. Esta es una limitación que, con los datos disponibles, no es posible solventar.

2.1.4. Análisis descriptivo de la variable dependiente generación de residuos

En este último apartado sobre la variable dependiente, se realiza un análisis descriptivo, contextualizando tanto la distribución como la evolución de la generación de residuos a nivel de distrito.

En el periodo completo analizado (2018-2023), se cuenta con un total de 126 observaciones, es decir, los residuos generados en cada distrito (21) en cada uno de los diferentes años (6).

El histograma contenido en la Figura 1 muestra la distribución de las 126 observaciones, que siguen aproximadamente una distribución normal. En todo el periodo, la generación promedio de residuos de un distrito fue de 55.870 toneladas. Aunque, como puede apreciarse, existe una considerable variabilidad, con observaciones por debajo de las 20.000 toneladas mientras otras acercándose a las 100.000 toneladas.



Fuente: Elaboración propia a partir de datos del Ayuntamiento de Madrid

Analizando los datos correspondientes a cada ejercicio, se observa que la generación total de residuos ha superado el millón de toneladas anuales de forma sistemática, aunque, en general, se ha ido reduciendo con el transcurso de los años. En 2023 se produce un incremento que se puede deber a la incorporación de datos correspondientes a grandes productores, que no se encuentran disponibles para ejercicios anteriores.

Tabla 4. Estadísticos de resumen de la generación de residuos por distrito y año

Concepto	2018	2019	2020	2021	2022	2023
Totales	1.224.446	1.212.731	1.130.257	1.156.157	1.141.244	1.174.824
Promedio	58.307	57.749	53.822	55.055	54.345	55.944
Mínimo	22.741	21.792	17.627	18.201	18.982	21.781
Máximo	90.873	91.385	90.241	84.945	83.815	96.424

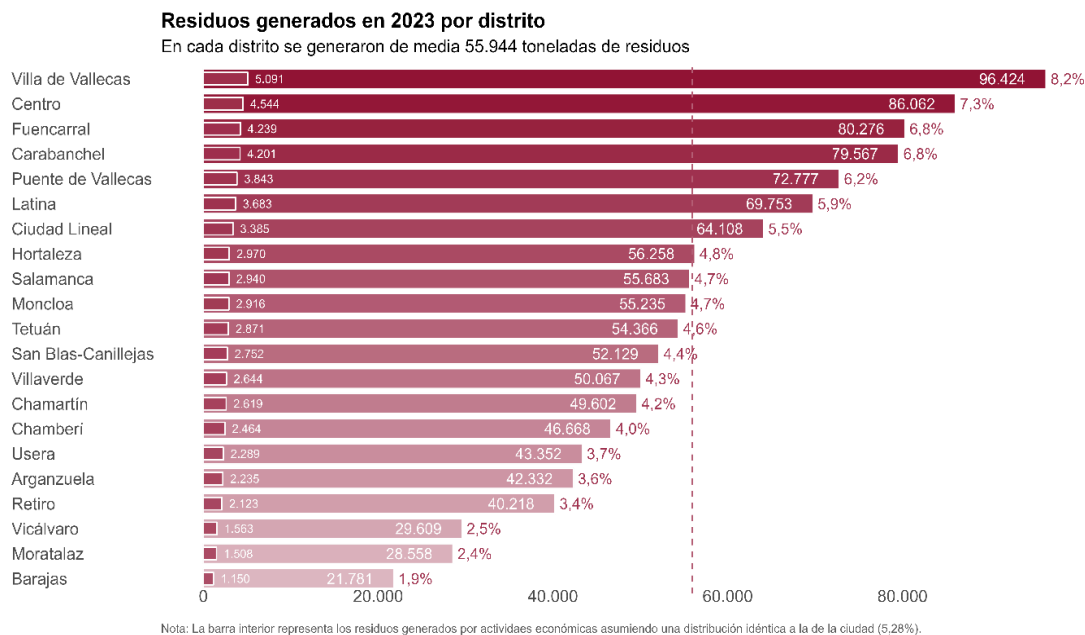
Desv.Est	19.626	19.137	19.261	18.996	18.463	19.515
----------	--------	--------	--------	--------	--------	--------

Fuente: Elaboración propia a partir de datos del Ayuntamiento de Madrid

La cantidad de residuos promedio por distrito ha ido oscilando, aunque replica el patrón del total. En 2020, año de pandemia, se obtuvo la menor generación total y promedio por distritos de todo el periodo. Además, la variabilidad entre los distritos se ha mantenido estable con una desviación estándar en torno a las 19.000 toneladas.

En cuanto a la variabilidad de la generación por distritos, destaca la amplitud de esta en 2023, como se puede observar en la figura 2. En ese sentido, Villa de Vallecas, Centro y Fuencarral superaron las 80.000 toneladas. Por otro lado, Vicálvaro, Moratalaz y Barajas se encontraron por debajo de las 30.000 toneladas.

Figura 2. La generación de residuos por distrito en 2023



Fuente:

Elaboración propia a partir de datos del Ayuntamiento de Madrid

La evolución temporal de la variable (Figura 3) muestra que, en la mayoría de los distritos, se aprecia una tendencia hacia la baja en la generación de residuos, principalmente durante el 2022, año de la Covid-19. En todo caso, se observan características diferenciadoras entre ellos:

- **Villa de Vallecas** es el distrito con **mayor generación de residuos** en 3 de los 6 años analizados, probablemente causado por Mercamadrid, un gran generador de residuos.
- Los distritos que se han mantenido durante el periodo como los generadores de menor número de toneladas son **Vicálvaro, Moratalaz y Barajas**. Estos distritos cuentan con

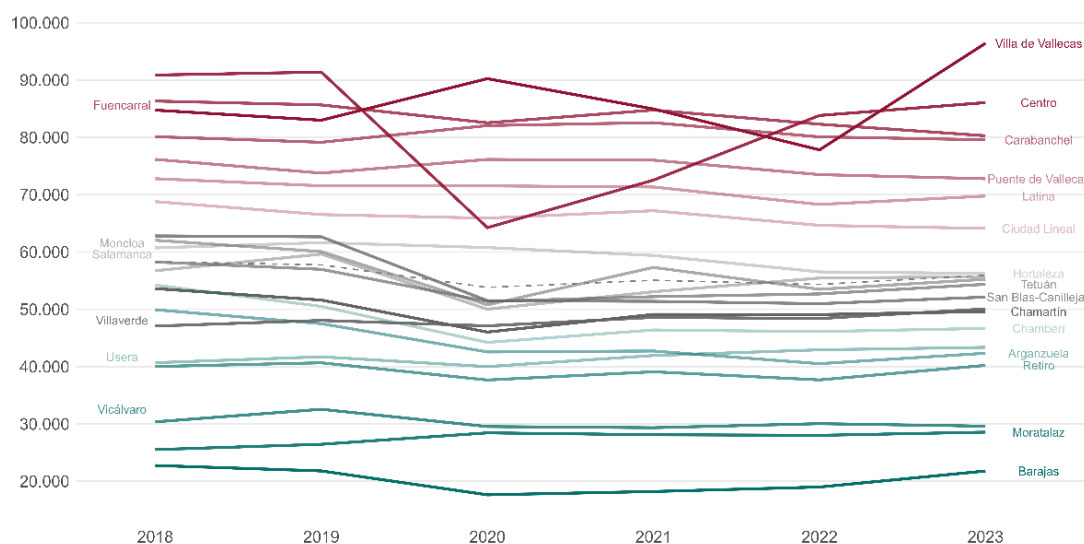
los porcentajes más bajos de población y actividad económica del conjunto de la Ciudad de Madrid.

- El distrito que se vio **más afectado por la pandemia fue el distrito Centro**, donde se produjo una caída considerable en 2020 que **no se visualiza en ningún otro distrito**. Esto se debe a que el tejido económico de este distrito recae principalmente en el sector servicios, en particular hostelería y restauración, los sectores más afectados por la pandemia.
- **Entre los 5 mayores generadores de residuos, son 4 los distritos que presentan una reducción de las toneladas** en el período estudiado: **Centro, Carabanchel, Puente de Vallecas y Latina**. Sin embargo, son reducciones muy pequeñas, de en torno al 3%, que, de nuevo, se pueden atribuir al efecto pandémico.

Figura 3. La generación de residuos a lo largo del periodo 2018-2023

Residuos generados a lo largo del periodo 2018-2023

El promedio anual por distrito ha oscilado entre los 50.000 y 60.000 toneladas



Fuente: Elaboración propia a partir de datos del Ayuntamiento de Madrid

2.2. Las variables independientes del modelo: la realidad sociodemográfica y económica de los distritos

El análisis de la generación de residuos en la Ciudad de Madrid ha puesto de manifiesto dos hechos relevantes a tener en cuenta en el proceso de identificación de zonas homogéneas:

1. La generación de residuos goza de una **gran disparidad entre distritos y en su trayectoria temporal**
2. **Los residuos son generados en una proporción muy superior por los hogares** que por las actividades económicas (Tabla 3).

Asimismo, es preciso destacar que la generación de residuos por parte de las actividades económicas no es homogénea entre los distritos. En este sentido, cabe señalar nuevamente el especial impacto de los grandes productores, especialmente Mercamadrid, en aquellos distritos en los que se ubican, implicando una generación más elevada.

Al no disponer de la información sobre las toneladas generadas por las actividades económicas en cada distrito, así como ser necesario identificar el mayor peso de la generación por hogar, es preciso reconocer variables exógenas (independientes) que permitan capturar las diferentes relaciones entre la generación de residuos por distritos. Por ello, se pone el foco en aspectos socioeconómicos, demográficos y de actividades económicas relativas a dichos distritos.

En la Memoria Técnica se presentó una propuesta inicial de 11 variables relacionadas con la estructura sociodemográfica y socioeconómica de los distritos que podrían determinar la generación de residuos (Tabla 5).

Tabla 5. Variables independientes inicialmente propuestas

	Ámbito	Indicador	Fuente	Unidad territorial
1	Estructura demográfica	<i>Densidad de población</i>	Estadística: Distritos en cifras (información de barrios)	Distrito y Barrio
2		<i>% de personas <35 años</i> <i>% de personas >65 años</i>	Estadística: Distritos en cifras (información de barrios)	Distrito y Barrio
3		<i>Índice de envejecimiento</i>	Banco de Datos del Ayuntamiento de Madrid: Áreas de Información Estadística	Distrito y Barrio
4	Nivel socioeconómico	<i>Renta neta media por hogar</i>	INE: indicadores de renta media y mediana	Distrito y Barrio
5		<i>Tasa de paro registrado</i>	Estadística: Distritos en cifras (información de barrios)	Distrito y Barrio
6	Diversidad cultural	<i>Porcentaje de población extranjera</i>	Estadística: Distritos en cifras (información de barrios)	Distrito y Barrio
7			Banco de Datos del Ayuntamiento de Madrid: Población clasificada por nacionalidad	Distrito y Barrio

	Ámbito	Indicador	Fuente	Unidad territorial
8	Actividades económicas	<i>Porcentaje de locales de hostelería</i>	Estadística: Distritos en cifras (información de barrios)	Distrito y Barrio
9	Afluencia turística	<i>Porcentaje de locales de servicios de alojamiento</i>	Estadística: Distritos en cifras (información de barrios)	Distrito y Barrio
10	Mercado de la vivienda	<i>Precio medio de la vivienda</i>	Banco de Datos del Ayuntamiento de Madrid: Precio medio declarado de la vivienda	Distrito y Barrio
11	Residuos generados	<i>Toneladas de residuos por fracción</i>	Portal de datos abiertos del Ayuntamiento: Recogida de residuos	Distrito

Fuente: elaboración propia.

Sin embargo, la evaluación de los datos disponibles y las nuevas sugerencias aportadas por el Ayuntamiento para tratar de captar la mayor cantidad de diferencias en generación posibles ha derivado en un conjunto de variables diferente y que se presenta y explica a continuación.

2.2.1. La estructura demográfica y económica de los distritos

El estudio demográfico y económico de los distritos ha derivado en el análisis de cuatro variables:

- Densidad de población
- Índice de envejecimiento
- Renta neta media por hogar
- Precio medio de la vivienda

Inicialmente, tal y como se recoge en la tabla anterior, se propuso el análisis de otras variables como: el porcentaje de personas menores de 35 años y mayores de 65 años, el porcentaje de población extranjera y la tasa de paro registrada. Sin embargo, tuvieron que descartarse debido a que los primeros modelos creados determinaban que no eran relevantes en la generación de residuos.

Tabla 6. Indicadores demográficos y económicos

Ámbito	Indicador	Fuente	Unidad territorial	Intervalo temporal
Estructura demográfica	Densidad de población	Portal de Datos del Ayto.	Distrito	2018-2023
	Índice de envejecimiento	Banco de Datos del Ayto.	Distrito	2018-2023

Estructura socioeconómica	Renta neta media por hogar	Atlas de distribución de renta de los hogares (INE).	Distrito	2018-2021
	Precio medio de la vivienda (€/m2)	Banco de Datos del Ayto.	Distrito	2018-2022

Fuente: Elaboración propia

De ese modo, a partir de los datos publicados en el Portal de datos, el Banco de Datos del Ayuntamiento y el Instituto Nacional de Estadística (INE) se ha construido un panel con información a nivel de distrito desde 2018 hasta 2023.

En la formación de dicho panel se han producido **dos limitaciones**:

- La información de renta neta media del hogar no contiene la información correspondiente a los años 2022 y 2023
- El precio medio de la vivienda no está disponible para el ejercicio 2023

Ante dichas limitaciones, se ha procedido a utilizar la información de los ejercicios previos, y se ha optado por regresar la variable correspondiente sobre la variable de año, imputándose los valores faltantes mediante una predicción lineal.

En la siguiente tabla se muestran los promedios de cada variable estudiada.

Tabla 7. Promedios de los indicadores a nivel de distrito

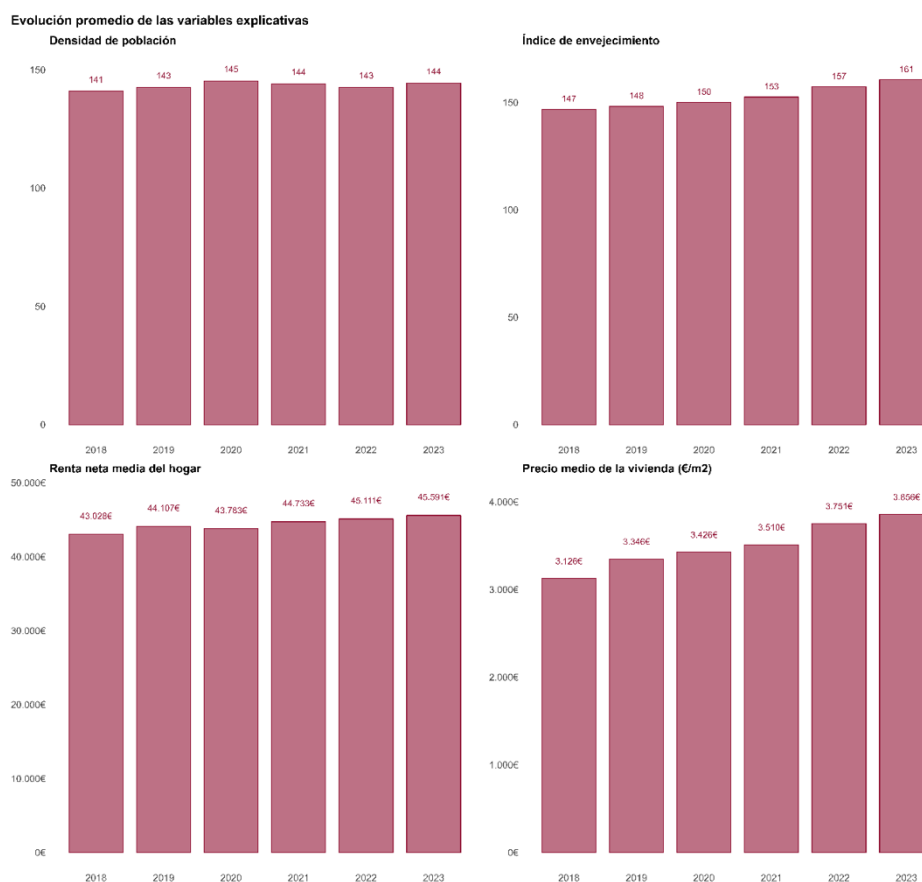
Variable	2018	2019	2020	2021	2022	2023
Densidad de población	140,9	142,5	145,3	144,0	142,6	144,3
Índice de envejecimiento	146,8	148,2	150,1	152,5	157,3	160,5
Precio medio de la vivienda (€/m2)	3.126 €	3.346 €	3.426 €	3.510 €	3.751 €	3.856 €
Renta neta media del hogar	43.028 €	44.107 €	43.783 €	44.733 €	45.111 €	45.591 €

Fuente: Elaboración propia a partir de datos del Ayuntamiento de Madrid

Como puede apreciarse, en todas las variables se ha producido un incremento de su valor a lo largo del periodo, siendo mayor en algunas variables. Comparando 2018 con 2023, la densidad de población aumentó solamente un 2,4% mientras que la renta neta media por hogar lo hizo en un 6%.

Los gráficos siguientes (Figura 4) muestran más claramente el desglose de esta evolución por las variables independientes. Como se observa, la densidad de población muestra este incremento mínimo.

Figura 4. Evolución de las variables demográficas y económicas



Fuente: Elaboración propia a partir de datos del Ayuntamiento de Madrid y del INE

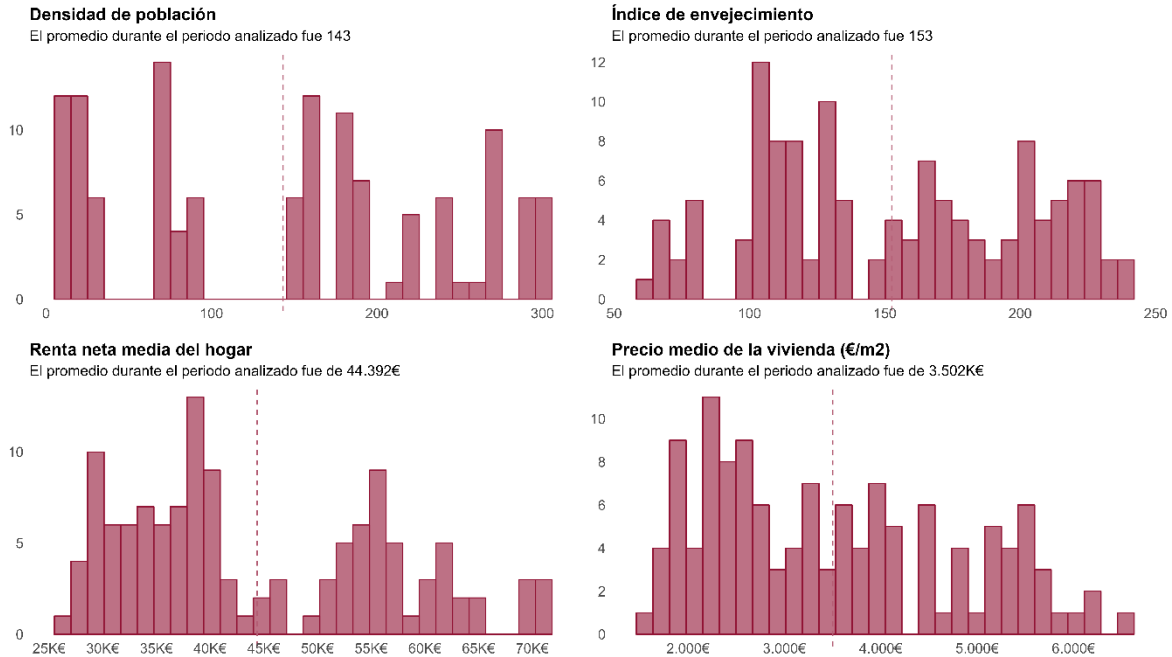
Por el contrario, los mayores aumentos han ocurrido en el precio medio de la vivienda (23,4%) y el índice de envejecimiento (9,3%)⁴. Aunque no se muestre en este conjunto de gráficos, a nivel de distrito se observa el mismo patrón de comportamiento descrito.

Por otra parte, los gráficos de distribución de cada una de las variables, que se muestran a continuación (Figura 5), permiten observar que las distribuciones de estas son sustantivamente distintas: **la densidad de población está concentrada en torno a una serie de valores, mientras que la renta del hogar y el precio de la vivienda siguen una distribución asimétrica positiva**, con unos pocos casos extremos en la cola derecha.

Figura 5. Distribución de las variables demográficas y económicas

⁴ Los porcentajes de aumento se calcularon comparando los valores del 2018 con los del 2023.

Distribución a nivel de distrito (2018-2023)



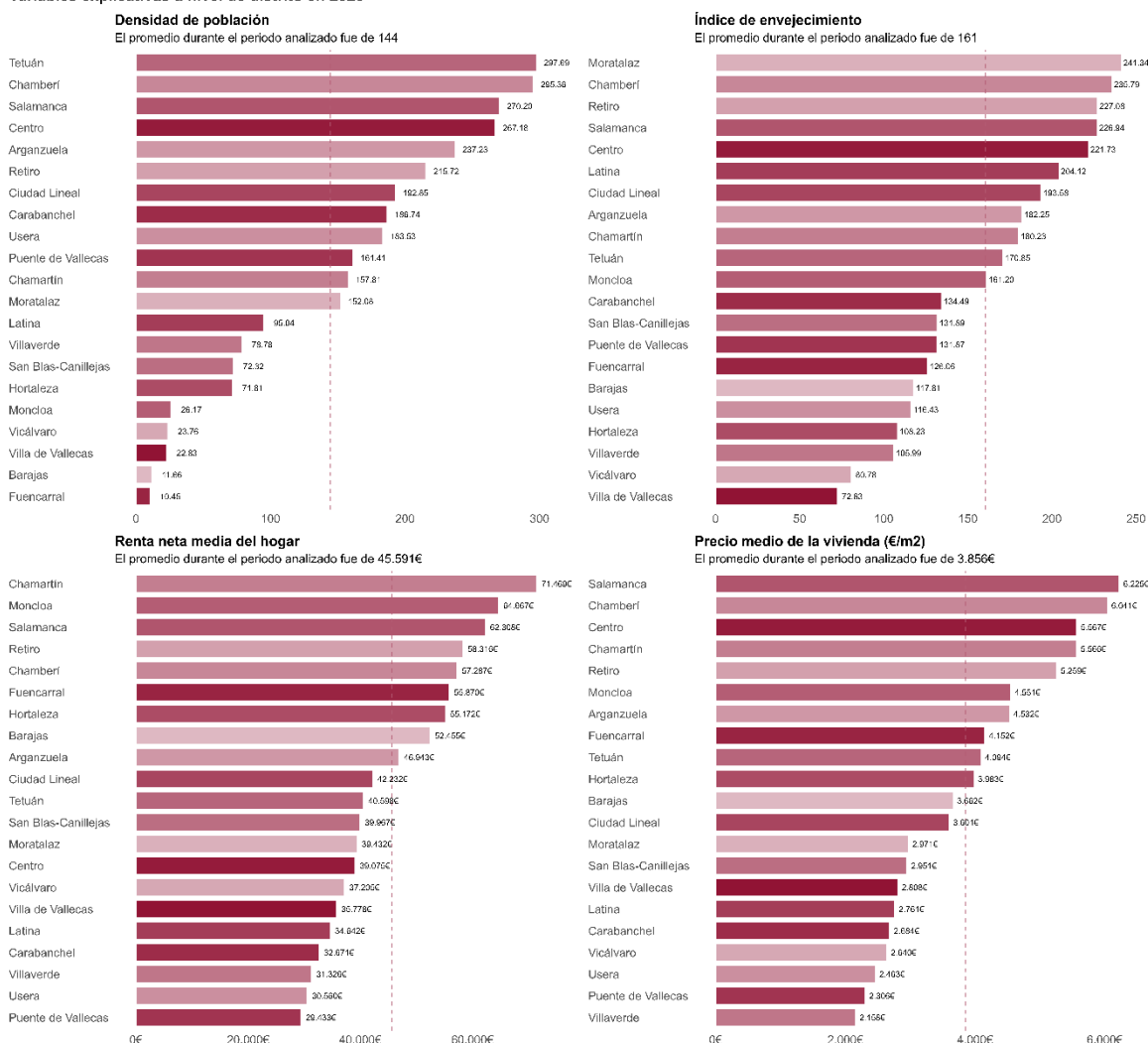
Fuente: Elaboración propia a partir de datos del Ayuntamiento de Madrid

Por último, en el conjunto de gráficos siguiente (Figura 6) se muestran las variables a nivel de distrito para 2023. Los aspectos observados más relevantes son los siguientes:

- Los **distritos que muestran los valores más altos para cada una de las variables son siempre distintos**. Es decir, ningún distrito se encuentra entre los cinco primeros siempre.
- Los distritos como **Chamartín y Salamanca** están en las **primeras posiciones** en términos de **densidad, índice de envejecimiento, renta del hogar y precio de la vivienda**.
- Los distritos **de Villaverde, Usera y Puente de Vallecas** muestran valores promedio de densidad y envejecimiento, pero son los que presentan menor renta y precios de la vivienda más bajos.

Figura 6. Las variables demográficas y económicas en 2023 por distrito

Variables explicativas a nivel de distrito en 2023



Fuente: Elaboración propia a partir de datos del Ayuntamiento de Madrid

La descripción de las variables demográficas y económicas muestran la heterogeneidad de los distritos. No obstante, parecen existir algunas relaciones lógicas entre las variables, que desembocan en las siguientes conclusiones:

- Los distritos de **Salamanca, Chamartín, Chamberí y Centro** son los distritos con **mayor poder adquisitivos**: su renta neta media por hogar y su precio de la vivienda puntúan valores altos. Es razonable encontrar pues que es en estos distritos donde se observan también los mayores índices de envejecimiento. Las personas más mayores, que generalmente tienen rentas más altas, residen en estos distritos.
- En el extremo opuesto, **Puente de Vallecas, Usera y Villaverde** son los distritos con **niveles adquisitivos más bajos**: su renta neta por hogar y precio de la vivienda arrojan valores más reducidos. Por ello, no es sorprendente encontrar a personas más jóvenes

con rentas y empleos menos estables residiendo aquí, dando lugar a índices de envejecimiento más bajos.

Con respecto a la densidad de población, no se observa un patrón fijo. Mientras que los distritos más ricos generalmente tienen densidades de población más elevadas, los distritos más empobrecidos se posicionan en el promedio. Esto último se puede deber al mayor número de convivientes en las viviendas propias de los barrios y distritos con menor nivel adquisitivo.

2.2.2. Actividades económicas: registro de locales

Las tareas econométricas y de clusterización desarrolladas condujeron a la **decisión de ampliar el análisis, introduciendo nuevas variables vinculadas a las actividades económicas**. El objetivo perseguido era disponer de indicadores que permitieran establecer agrupaciones de distritos asignando mayor peso a factores vinculados a las actividades.

De ese modo, se han extraído, del Portal de Datos:

- La estadística de **infraestructura hotelera**, con datos disponibles desde 2018 hasta 2022 a nivel de distrito. Esta estadística proporciona el total de establecimientos de alojamiento: hoteles, hostales y pensiones.
- El **Censo de locales y actividades**, que registra el número de actividades en locales abiertos con acceso a puerta de calle. El número de locales está desagregado para cada distrito y barrio en función de la actividad económica (CNAE 1 dígito), como se observa en la Tabla 8.

Tabla 8. Clasificación CNAE

Código CNAE	Actividad
A	Agricultura, ganadería, silvicultura y pesca
B	Industrias extractivas
C	Industria manufacturera
D	Suministro de energía eléctrica, gas, vapor y aire acondicionado
E	Suministro de agua, actividades de saneamiento, gestión de residuos y descontaminación
F	Construcción
G	Comercio al por mayor y al por menor; reparación de vehículos de motor y motocicletas
H	Transporte y almacenamiento
I	Hostelería
J	Información y comunicaciones
K	Actividades financieras y de seguros
L	Actividades inmobiliarias
M	Actividades profesionales, científicas y técnicas
N	Actividades administrativas y servicios auxiliares

Código CNAE	Actividad
O	Administración Pública y defensa; Seguridad Social obligatoria
P	Educación
Q	Actividades sanitarias y de servicios sociales
R	Actividades artísticas, recreativas y de entretenimiento
S	Otros servicios
T	Actividades de los hogares como empleadores de personal doméstico
U	Actividades de organizaciones y organismos extraterritoriales

Fuente: Elaboración propia

La información del Censo está disponible para todo el periodo de interés (2018-2023) y se facilita para dos momentos de cada año: en enero y en julio. Los datos seleccionados han sido los correspondientes a julio debido del que se disponen datos en 2023. Por ello, es el dato más reciente. Manteniendo el foco en julio en los anteriores años se garantiza la comparabilidad del dato y se evitan variaciones estacionales.

Por otra parte, a nivel operativo se han agrupado las actividades de la forma que se indica en la Tabla 8. Dado que no se dispone de información sobre la infraestructura hotelera para el año 2023, sus valores se imputan aprovechando la información de los años previos, replicando la estrategia seguida para las variables de renta y precio de la vivienda.

Una vez realizadas las tareas de limpieza y transformación de los datos, se ha generado un nuevo panel a nivel de distrito que se combina con la información sociodemográfica.

Tabla 9. Indicadores de actividades económicas

Ámbito	Indicador	Descripción
Infraestructura hotelera	Infraestructura hotelera	N.º de hoteles, hostales y pensiones
Actividades económicas	Agricultura	N.º de locales de Agricultura (A)
	Industria	N.º de locales de la Industria Extractiva (B) y Manufacturera (C)
	Construcción	N.º de locales de la Construcción (F)
	Servicios	N.º de locales de Energía (D), Agua (E), Comunicaciones (J), Finanzas (K), Inmobiliarias (L), Profesionales (M), Administrativas (N), Administración Pública (O), Educación (P) y Sanidad (Q)
	Comercio	N.º de locales de Comercio (G)
	Hostelería	Nº de locales de Hostelería (I)
	Entretenimiento	N.º de locales de Entretenimiento (R)

Fuente: Elaboración propia

En la siguiente tabla (Tabla 9) se muestra el número de locales promedio en los distritos para cada año y para cada uno de los indicadores, junto con el indicador de infraestructura hotelera. Las características más relevantes observadas en la información que proporcionan los promedios son las siguientes:

- El número de actividades en Agricultura es inferior a 1.
- El número de locales promedio vinculados a Industria, Construcción y Comercio ha disminuido durante el periodo, mientras que en Hostelería, Entretenimiento y Servicios ha aumentado.
- Existe una elevada variación en el número de locales según el tipo de actividad.
- A pesar del ascenso del número de locales asociados a Hostelería, el correspondiente a Infraestructura hotelera se ha reducido ligeramente a lo largo del periodo.

Tabla 10. Promedio de los indicadores de actividades económicas por año

Actividades	2018	2019	2020	2021	2022	2023
Agricultura	0,6	0,6	0,6	0,6	0,4	0,4
Construcción	139	137	136	135	113	113
Comercio	1.943	1.931	1.917	1.912	1.804	1.790
Hostelería	926	936	948	958	933	936
Industria	142	147	151	151	133	133
Entretenimiento	134	139	141	136	141	141
Servicios	990	1.018	1.042	1.052	1.045	1.043
Infraestructura hotelera	50	50	45	46	48	46

Fuente: Elaboración propia a partir de datos del Ayuntamiento de Madrid

En el conjunto de la Ciudad de Madrid, **el número de locales más elevado se encuentra en la actividad Comercio**, como se puede observar en la tabla siguiente, **seguido por Servicios y Hostelería** a una distancia relevante.

Entre 2018 y 2023 han aumentado los locales asociados a tres actividades concretas: Hostelería, Entretenimiento y Servicios.

Tabla 11. Total de los indicadores de actividades económicas por año

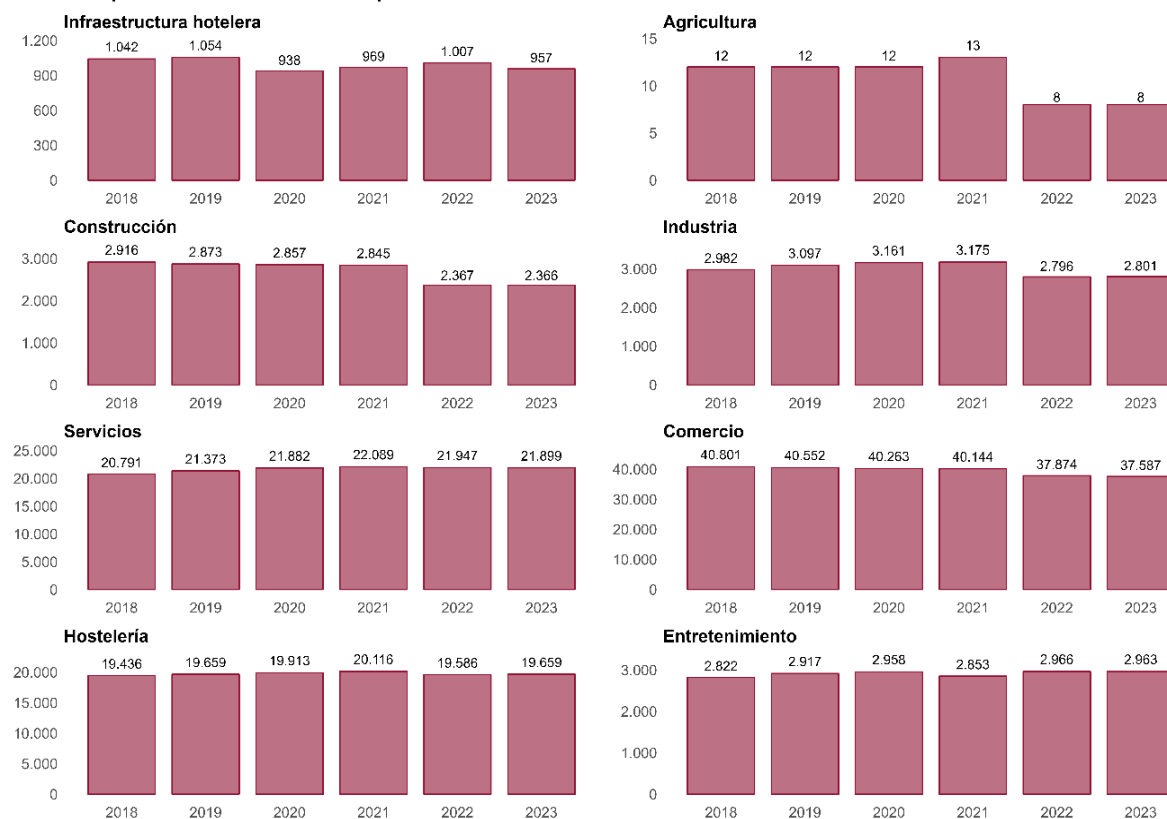
Actividades	2018	2019	2020	2021	2022	2023
Agricultura	12	12	12	13	8	8
Construcción	2.916	2.873	2.857	2.845	2.367	2.366
Comercio	40.801	40.552	40.263	40.144	37.874	37.587
Hostelería	19.436	19.659	19.913	20.116	19.586	19.659
Industria	2.982	3.097	3.161	3.175	2.796	2.801
Entretenimiento	2.822	2.917	2.958	2.853	2.966	2.963
Servicios	20.791	21.373	21.882	22.089	21.947	21.899
Infraestructura hotelera	1.042	1.054	938	969	1.007	957

Fuente: Elaboración propia a partir de datos del Ayuntamiento de Madrid

En los gráficos siguientes se observa de forma más visual esta evolución.

Figura 7. Evolución del número de locales en cada actividad 2018-2023

Evolución promedio de las variables explicativas



Fuente: Elaboración propia a partir de datos del Ayuntamiento de Madrid

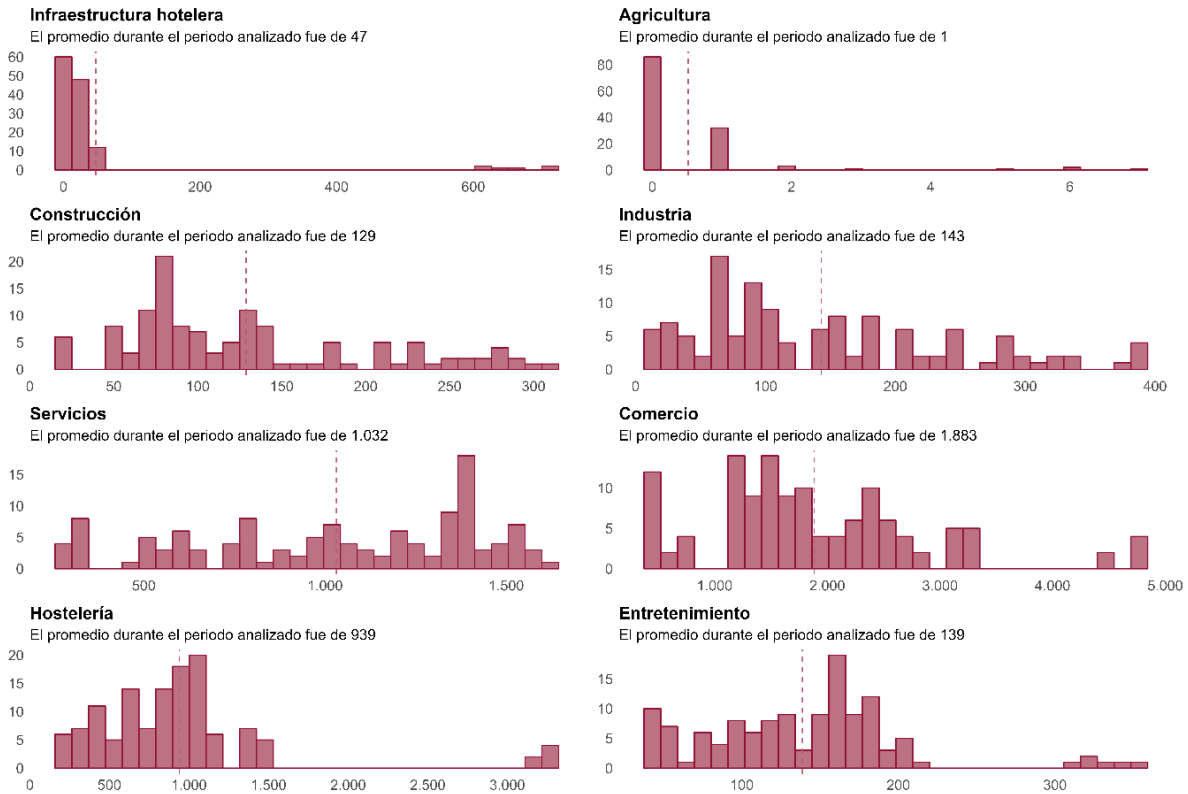
En estos gráficos (Figura 7) se muestra la caída de la infraestructura hotelera, principalmente a partir de los años de la pandemia. Se observan, no obstante, también atisbos de recuperación en el 2022. También, se comprueban las reducciones más significativas en los locales de industria y construcción, ambas a partir del 2021.

A excepción de los locales de entretenimiento y servicios, todas las actividades económicas parecen estar afectadas por el efecto pandemia (2020) o postpandemia (2021-2022).

La distribución de los indicadores de actividades económicas (Figura 8), muestra en la mayoría de los casos el mismo patrón que en los determinantes socioeconómicos (Figura 5): las distribuciones presentan asimetría positiva con unos pocos casos en el extremo derecho de la distribución.

Figura 8. La distribución de los indicadores de actividades económicas en 2023

Distribución a nivel de distrito (2018-2023)



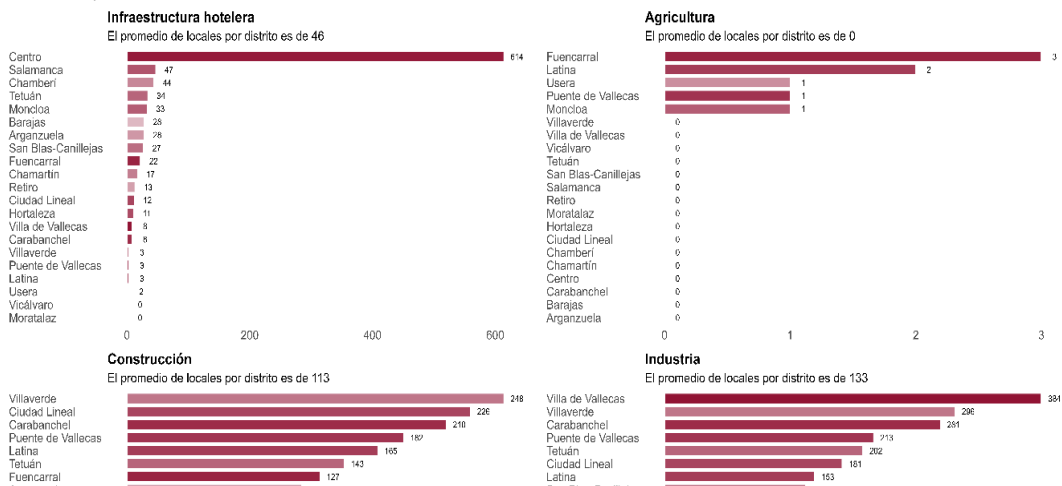
Fuente: Elaboración propia a partir de datos del Ayuntamiento de Madrid

Finalmente, el último conjunto de gráficos (Figura 8) presenta los datos para cada distrito y para el ejercicio 2023. Destaca el caso del distrito **Centro**, que presenta un **número de locales en actividades de hostelería, entretenimiento e infraestructura hotelera** sensiblemente superior al del resto. Por el otro lado, Vicálvaro se sitúa en las últimas posiciones en buena parte de los indicadores.

Asimismo, se aprecia que coinciden en las primeras posiciones algunos distritos en indicadores como en Comercio, Hostelería y Entretenimiento. Estos son los distritos con mayores niveles adquisitivos, Centro, Salamanca y Chamberí, y, como *outlier*, Carabanchel.

Figura 9. Las variables de actividades económicas por distrito en 2023

Variables explicativas a nivel de distrito en 2023



Fuente: Elaboración propia a partir de datos del Ayuntamiento de Madrid

2.2.3. La influencia del sector turístico en la generación de residuos

El Ayuntamiento de Madrid propuso ampliar el análisis de las actividades económicas con otro específico sobre el **sector turístico y su relación con la generación de residuos**, tomando como base el número de puntos de interés turístico presentes en cada distrito.

Con ese fin, el Ayuntamiento facilitó una base de datos con los puntos de interés de cada distrito en 2023. Para el propósito de este estudio, de dicho listado se seleccionaron los puntos clasificados como: edificios y monumentos, instalaciones culturales, instalaciones deportivas, parques y centro de ocio y parques y jardines.

El listado completo constaba de **886** regístritos, de los cuales 583 contaban el nombre del punto de interés, la dirección, la ubicación (latitud y longitud) y el distrito. Para los 303 registros

restantes se desconocía el distrito, pero se contaba con la ubicación, y mediante técnicas de geolocalización se les pudo asignar el correspondiente distrito.

En la siguiente tabla se muestra el número de ubicaciones de interés turístico por categoría y para cada uno de los distritos. Destaca el distrito Centro con 345 puntos de interés turístico. En la cola de la distribución se ubica Moratalaz con únicamente 1.

El número de puntos de interés turístico de media es de **37 por distrito**, con 11 de mediana, lo que **denota una marcada distribución asimétrica**, con unos pocos distritos que concentran la mayoría de las ubicaciones de interés.

Tabla 12. Puntos de interés turístico por distrito (2023)

Distrito	TOTAL	Edificios y monumentos	Instalaciones culturales	Instalaciones deportivas	Parques y jardines	Parques y centros de ocio
Centro	345	153	178	0	14	0
Chamberí	71	25	44	0	2	0
Moncloa-Aravaca	59	27	18	1	10	3
Arganzuela	55	10	38	2	3	2
Salamanca	52	15	33	1	1	2
Retiro	50	22	22	0	5	1
Carabanchel	26	8	13	1	3	1
Chamartín	24	7	11	1	3	2
Fuencarral-El Pardo	17	11	3	0	2	1
Tetuán	12	8	3	0	0	1
San Blas-Canillejas	11	2	3	2	2	2
Latina	10	2	6	1	0	1
Puente de Vallecas	10	3	3	1	3	0
Barajas	7	3	1	0	3	0
Hortaleza	6	4	0	0	2	0
Usera	6	1	3	1	1	0
Ciudad Lineal	6	3	3	0	0	0
Villa de Vallecas	6	3	2	0	1	0
Vicálvaro	3	2	0	0	0	1
Villaverde	2	1	1	0	0	0
Moratalaz	1	0	0	0	1	0

Fuente: Ayuntamiento de Madrid

2.3. Determinantes de la generación de residuos de la Ciudad de Madrid por distrito

Una vez construido el panel de datos a nivel de distrito y analizadas descriptivamente las variables, la siguiente tarea consiste en **aplicar los modelos de regresión con el fin de identificar las variables relevantes.**

La propuesta inicial de trabajo contemplaba el objetivo de construir un panel a nivel de distrito y barrio utilizando como variable dependiente la generación de residuos con información correspondiente a datos estadísticos. Sin embargo, el desarrollo de las tareas ha conducido a considerar únicamente la generación por distritos y la información administrativa de las cantidades asociadas a esta generación, como anteriormente se señalaba⁵.

La utilización de la información sobre generación de residuos **estadística u administrativa** requería analizar las características de ambas fuentes de datos, y la mayor o menor idoneidad de una u otra dependía de los objetivos perseguidos con los ejercicios a realizar. Por un lado, **la información estadística captura conjuntamente la generación de residuos generados por hogares y por actividades económicas.** De esta forma permite identificar mejor qué variables socioeconómicas son relevantes y, con ello, construir posteriormente clústers homogéneos en términos socioeconómicos. Por el otro, **la información administrativa recoge más adecuadamente los residuos generados por las actividades económicas.** De este modo, las variables explicativas de las actividades económicas deberían dar resultados más significativos y, posteriormente, generar unos clústers más homogéneos a nivel de actividades –aunque no necesariamente a nivel geográfico ni socioeconómico-.

En cualquier caso, se han construido modelos de corte transversal y de panel considerando ambas fuentes de datos de la variable dependiente, con el fin de disponer de las conclusiones más completas posibles.

A continuación, se exponen los resultados obtenidos del análisis de:

- El modelo de datos de corte transversal 2023
- El modelo de datos panel 2018-2023

La variable de puntos de interés turístico no se ha podido utilizar en el modelo de datos de panel al tratarse de una constante. Su efecto estaría capturado precisamente por la constante del modelo. Si está incluida, sin embargo, en el modelo de datos transversales de 2023.

⁵ No están disponibles los datos de la generación por barrios para todo el período estudiado en el trabajo y la información administrativa sobre la variable es más completa para los objetivos perseguidos que la estadística.

2.3.1. Modelo de datos de corte transversal 2023

En primer lugar, se muestran los resultados para el modelo con datos de corte transversal del año 2023, incluyendo todas las variables explicativas analizadas en el apartado anterior (sociodemográficas, actividades económicas y puntos de interés turístico).

En la tabla siguiente (Tabla 11) se muestran los resultados de cuatro modelos, dos utilizando como variable dependiente los residuos “administrativos” y, en los otros dos, la variable “estadística”. En ambos casos se prueba un modelo con todas las variables explicativas y un segundo modelo eliminando algunas de las no significativas.

Los hallazgos más relevantes son los siguientes:

- **Los modelos completos (independientemente de la variable dependiente) poseen un mejor ajuste** (R2 ajustado, AIC y BIC) que los modelos reducidos. Por tanto, más allá de la significatividad individual de las variables, con el objetivo de desarrollar el análisis de clúster, parece razonable **incorporar todas las variables utilizadas**.
- Únicamente la variable de actividades económicas vinculadas a la **Industria posee un resultado significativo** (y positivo) **en todos los modelos**, es decir, tanto en la generación de residuos, con datos administrativos como estadísticos.
- El número de actividades de **Agricultura, Hostelería** (Columna 1) e **Industria**, muestran una **relación significativa con la generación de residuos**.
- Al utilizar como variable dependiente la variable de residuos “estadística”, también son significativas la **densidad de población y las actividades vinculadas a los Servicios** (columna 4).

Tabla 12. Resultados del modelo con datos transversales para 2023

	VD: Residuos (Administrativo)	VD: Residuos (Administrativo)	VD: Residuos (Estadístico)	VD: Residuos (Estadístico)
(Intercept)	48971.4 (235611.1)	-103554.1 (120605.7)	66819.3 (182566.4)	177033.2 (107064.3)
Densidad de población	-25.4 (83.6)	-31.6 (26.0)	-13.5 (64.4)	-57.7* (26.0)
Índice de envejecimiento	17.5 (55.7)		17.2 (50.3)	
log Renta neta media del hogar	26396.9 (57041.6)	11506.1 (11347.2)	34094.0 (47586.7)	-15163.6 (10144.4)

	VD: Residuos (Administrativo)	VD: Residuos (Administrativo)	VD: Residuos (Estadístico)	VD: Residuos (Estadístico)
log Precio medio de la vivienda (€/m2)	-39055.7 (53585.0)		-52463.9 (45820.5)	
Infraestructura hotelera	9.4 (79.7)	13.6 (69.2)	-25.9 (66.8)	-14.4 (70.7)
Agricultura	7100.1 (3235.6)	8234.4** (2025.6)	3711.2 (2620.4)	4425.3 (2170.8)
Construcción	-200.5 (111.1)		-87.7 (66.7)	
Industria	167.2* (48.8)	162.7** (43.4)	62.8* (24.6)	55.5** (16.6)
Servicios	31.4 (22.3)	13.4 (8.9)	16.8 (20.8)	24.5* (10.6)
Comercio	10.2 (9.1)		6.3 (6.9)	
Hostelería	-39.2* (15.9)	-2.0 (12.9)	-18.1 (14.9)	10.3 (14.8)
Entretenimiento	151.8 (198.5)		247.8 (145.9)	
Puntos de interés turístico	199.7 (138.2)	94.2 (103.4)	125.9 (105.6)	46.2 (114.7)
Observaciones	21	21	21	21
R2	0.946	0.884	0.957	0.907
R2 Adj.	0.846	0.806	0.876	0.845
AIC	440.2	446.3	429.8	435.8
BIC	454.8	455.7	444.4	445.2
Errores estándares	Robustos	Robustos	Robustos	Robustos

* p < 0.05, ** p < 0.01, *** p < 0.001. Nota: Regresión con datos transversales de distrito (21) y año (2023) y errores estándares robustos a la heterocedasticidad.

2.3.2. Modelo de datos panel 2018- 2023

El estudio de las relaciones de la variable dependiente con las candidatas a independientes se amplía con el asociado al desarrollo de modelos de datos de panel del período comprendido entre 2018 y 2023.

Los resultados se muestran siguiendo una lógica secuencial de modelos anidados:

1. Utilizando **variables socioeconómicas y demográficas**
2. Considerando variables de **actividades económicas y la influencia del sector turístico**
3. Finalmente, recogiendo **todas las posibles variables explicativas**

En el modelo panel se ha introducido como variable de control una *dummy* con valor 1 para el año 2020, con el objetivo de capturar el efecto producido por la pandemia del Covid-19.

Como se explicaba anteriormente, en el modelo panel no es posible introducir la variable de puntos de interés turísticos al tratarse de una constante en el tiempo.

Modelo dependiente de variables socioeconómicas y demográficas

En el caso de las variables sociodemográficas (Tabla 12), si se utiliza como variable dependiente los residuos provenientes tanto de la **información estadística** (columna 1) como administrativa (columna 2) para realizar un ejercicio de sensibilidad entre los resultados de las estimaciones de los modelos.

Como se puede observar en la Tabla 12, **las variables** que aparecen como **significativas** son: **la densidad de población, el índice de envejecimiento y la renta del hogar (en logaritmo), así como la variable ficticia que captura los efectos sobre las variables de la pandemia.**

La única diferencia entre ambos modelos (estadístico vs administrativo) es la significatividad de **la renta del hogar, que pasa a no ser diferente de cero para explicar la generación de residuos. Una posible explicación de este efecto causal puede ser la inclusión en la variable administrativo de una mayor variabilidad de los residuos generados por las actividades económicas** que no sean directamente atribuibles a los patrones de consumo de los hogares (ver discusión en [sección 2.1](#))

Tabla 13. Resultados de los modelos de regresión con datos de panel para las variables sociodemográficas

	VD: Residuos (Estadístico)	VD: Residuos (Administrativo)
Densidad de población	-459.6* (224.5)	-565.6* (264.3)
Índice de envejecimiento	-108.8* (46.8)	-212.3** (48.6)
log Renta neta media del hogar	61280.9** (23311.0)	34292.4 (25819.8)
log Precio medio de la vivienda (€/m2)	-6268.4 (5418.2)	-2177.4 (6440.5)
Covid (Dummy 1 = 2020)	-2539.4** (866.4)	-1196.2 (925.6)
Observaciones	126	126
R2	0.338	0.261
R2 Adj.	0.173	0.076
AIC	2349.0	2382.1
BIC	2366.0	2399.1
Errores estándares	Robustos-Clúster: Distrito	Robustos-Clúster: Distrito

* $p < 0.05$, ** $p < 0.01$. Nota: Regresión de efectos fijos con datos panel a nivel de distrito (21) y año (2018 a 2023) con efectos individuales y errores estándares robustos y clusterizados a nivel de distrito.

Aunque a priori parezca contraintuitiva, existe una relación negativa y significativa entre la **densidad de población y la variable de generación de residuos**. Una posible explicación de este factor se debería a que la densidad no mide la cantidad de población sino su concentración.

En el caso de los distritos de Madrid los distritos más densos, Tetuán, Salamanca, Arganzuela, Centro y Chamberí, no son los más poblados, sino que tiene una cantidad de población en torno a la media. Los aumentos en la densidad de población se asocian con reducciones en la generación de residuos debido a que este aumento captura la disminución del total de la población. Por su parte, los distritos más poblados poseen una densidad de población alrededor del promedio. El signo negativo está reflejando este patrón y capturando tanto la cantidad de población como su concentración.

Variables dependientes de actividades económicas y la influencia del sector turístico

En lo que respecta a las variables de actividades económicas, también se incluye un modelo con la variable dependiente estadística para continuar con la tónica presente de realizar un ejercicio de robustez de la estimación de los modelos.

En las primeras dos columnas (Tabla 14) se incluyen los resultados de todas las variables (para las distintas variables dependientes), mientras que en las tres siguientes se extraen las variables no significativas hasta obtener un modelo donde todas las variables son relevantes.

Comparando los resultados de la variable dependiente administrativa, el modelo con mejor ajuste es el de la cuarta columna, en el que se extraen algunas de las variables explicativas.

Como puede apreciarse, **la infraestructura hotelera y las actividades vinculadas a la Construcción se asocian con aumentos de la generación de residuos, mientras que las actividades vinculadas a los Servicios la disminuyen.**

En este caso, la variable *dummy* reflejando el año de la pandemia es significativa en todos los modelos, capturando el descenso de la generación de residuos que se produjo en 2020.

Tabla 14. Resultados de los modelos de regresión con datos de panel para las variables de actividades económicas

	VD: Residuos (Estadístico)	VD: Residuos (Administrativo)	VD: Residuos (Administrativo)	VD: Residuos (Administrativo)	VD: Residuos (Administrativo)
Infraestructura hotelera	141.5** (38.0)	155.6** (36.8)	157.8** (37.3)	166.3** (21.2)	147.9** (13.1)
Agricultura	520.6 (533.8)	656.4 (658.1)			
Construcción	86.8 (54.3)	100.6* (47.5)	112.6 (58.5)	106.0 (57.9)	42.3** (11.6)
Industria	5.4 (20.8)	16.2 (18.0)	15.8 (19.4)	16.7 (19.4)	
Servicios	-0.7 (16.5)	-40.7* (16.2)	-40.1** (14.1)	-45.7** (12.3)	-31.5** (11.0)
Comercio	-19.8 (14.6)	-17.7 (12.9)	-19.9 (15.1)	-20.7 (13.8)	
Hostelería	-13.2	-15.3	-13.4		

	VD: Residuos (Estadístico)	VD: Residuos (Administrativo)	VD: Residuos (Administrativo)	VD: Residuos (Administrativo)	VD: Residuos (Administrativo)
	(27.7)	(27.6)	(28.3)		
Entretenimiento	10.6	34.7			
	(71.3)	(80.3)			
Covid (Dummy 1 = 2020)	-3678.8**	-1711.4*	-1607.8*	-1569.0*	-1958.1*
	(848.8)	(743.3)	(628.1)	(607.3)	(847.1)
Observaciones	126	126	126	126	126
R2	0.468	0.449	0.442	0.436	0.372
R2 Adj.	0.307	0.282	0.288	0.288	0.223
AIC	2329.6	2353.1	2350.6	2350.0	2359.4
BIC	2357.9	2381.5	2373.3	2369.8	2373.6
Errores estándares	Robustos- Clúster: Distrito	Robustos- Clúster: Distrito	Robustos- Clúster: Distrito	Robustos- Clúster: Distrito	Robustos- Clúster: Distrito

* p < 0.05, ** p < 0.01. Nota: Regresión de efectos fijos con datos panel a nivel de distrito (21) y año (2018 a 2023) con efectos individuales y errores estándares robustos y clusterizados a nivel de distrito.

Los determinantes de la generación de residuos

Para finalizar el análisis de datos de panel, se construye un modelo conjunto con las variables socioeconómicas y de actividades económicas, cuyos resultados se pueden obtener en la Tabla 15:

- La primera columna muestra los resultados incluyendo todas las variables regresadas sobre la variable dependiente administrativa
- La segunda y tercera columnas recogen los resultados tras eliminar algunas de las variables no significativas (en estos modelos y en los anteriores).
- La cuarta y quinta columna replican los modelos, pero cambiando la variable dependiente, utilizando la variable estadística

En primer lugar, al igual que sucedió con los modelos transversales, **el modelo con mejor ajuste para ambas variables dependientes es el modelo completo donde se usan todas las variables explicativas, sin excluir a ninguna debido a su significatividad**, resultado que se debe tomar de interés para la posterior cauterización de los distritos de la ciudad de Madrid en zonas homogéneas (ver [sección 3.1.](#)).

En segundo lugar, **ninguna de las variables demográficas y económicas muestra una relación significativa con la variable generación de residuos, tanto en el modelo con datos administrativos como estadísticos**. Únicamente la renta neta del hogar es relevante a la hora de explicar la generación de residuos proveniente de datos estadísticos, donde el peso de los residuos generados por las actividades es menor, tal y como se ha expuesto anteriormente.

Tabla 15. Resultados de los modelos de regresión con datos de panel

	VD: Residuos (Administrativo)	VD: Residuos (Administrativo)	VD: Residuos (Administrativo)	VD: Residuos (Estadístico)	VD: Residuos (Estadístico)
Densidad de población	86.3 (144.6)	-76.6 (111.3)	-79.5 (109.1)	50.7 (187.7)	-112.1 (166.1)
Índice de envejecimiento	-36.2 (82.0)	-14.1 (126.2)	-17.6 (119.5)	-40.0 (99.1)	-27.0 (156.1)
log Renta neta media del hogar	19392.6 (17413.2)	30846.9 (17966.3)	28280.5 (14374.0)	36845.7* (15875.6)	25870.4* (11490.5)
log Precio medio de la vivienda (€/m2)	-3434.6 (8089.1)	-1244.8 (7462.6)		-11857.2 (8724.2)	
Infraestructura hotelera	164.4** (52.9)	138.7** (38.1)	137.2** (34.4)	153.8* (68.7)	114.8* (47.7)
Agricultura	669.1 (689.2)			519.2 (623.7)	
Construcción	110.4* (45.1)	56.3** (18.0)	57.4** (17.2)	88.7 (47.3)	24.6 (19.7)
Industria	11.9 (17.0)			-1.4 (17.5)	
Servicios	-42.3* (18.3)	-39.8** (14.4)	-40.1** (14.5)	2.0 (18.4)	1.0 (14.9)
Comercio	-19.2 (13.4)			-22.7 (14.6)	
Hostelería	-14.7			-10.7	

	VD: Residuos (Administrativo)	VD: Residuos (Administrativo)	VD: Residuos (Administrativo)	VD: Residuos (Estadístico)	VD: Residuos (Estadístico)
	(26.8)			(24.7)	
Entretenimiento	35.0			16.6	
	(92.2)			(82.6)	
Covid (Dummy 1 = 2020)	-1702.9	-1370.1	-1406.5	-3375.0**	-3474.3**
	(1000.6)	(848.3)	(906.9)	(1134.1)	(1100.7)
Observaciones	126	126	126	126	126
R2	0.459	0.385	0.385	0.500	0.391
R2 Adj.	0.265	0.208	0.216	0.321	0.223
AIC	2358.7	2364.8	2362.8	2329.6	2342.6
BIC	2398.4	2390.3	2385.5	2369.3	2365.3
Errores estándares	Robustos- Clúster: Distrito	Robustos- Clúster: Distrito	Robustos- Clúster: Distrito	Robustos- Clúster: Distrito	Robustos- Clúster: Distrito

* $p < 0.05$, ** $p < 0.01$. Nota: Regresión de efectos fijos con datos panel a nivel de distrito (21) y año (2018 a 2023) con efectos individuales y errores estándares robustos y clusterizados a nivel de distrito.

En este sentido, la **única variable que permanece como positiva y estadísticamente diferente de cero en todos los modelos** estimados es la referente al nivel de **infraestructura hotelera** por distrito, siendo una de las variables claves en explicar la generación de residuos del distrito.

La significatividad estadística de la renta del hogar sobre la generación de residuos en el modelo con variable estadística se ve **capturada por la significatividad de las variables referentes a la construcción y los servicios** del modelo con datos administrativos.

Mientras que la mayor presencia de las actividades dedicadas a la **construcción**, al igual que con mayor dotación de infraestructura hotelera, muestran una **relación positiva y significativa** con la generación de residuos, la mayor presencia de actividades dedicadas a los **servicios muestra una ligera relación negativa y significativa con la misma**.

Una posible explicación podría ser que los distritos con una **mayor concentración de locales de servicios tengan una población con hábitos de consumo diferentes**, lo que podría influir en la cantidad y el tipo de residuos generados. Por ejemplo, los distritos con una alta densidad de locales de servicios podrían tener una población con una **mayor conciencia ambiental**, lo que se traduce en una menor generación de residuos

Otra explicación podría estar relacionada con las características específicas de los establecimientos de servicios en esos distritos. Es posible que estos establecimientos estén implementando **prácticas de gestión de residuos más efectivas o estén utilizando materiales y envases que generan menos residuos en comparación con otros tipos de establecimientos**, como los afectados por la actividad de la construcción o el sector hotelero.

3. La identificación de zonas homogéneas en Madrid: un análisis clúster para la estratificación de los distritos en la ciudad

Como se ha visto, para comprender a fondo la dinámica de generación de residuos en la ciudad, se ha realizado un **análisis de inferencia de datos panel con efectos fijos**, abordando tres conjuntos clave de información: **variables de tipo socioeconómico y demográfico, variables relacionadas con la actividad económica y variables vinculadas a una mayor o menor afluencia del turismo**. Estos conjuntos de variables han sido seleccionados estratégicamente para capturar tanto la parte de generación de residuos inducida por los patrones de consumo de la población local – que reside en la ciudad de Madrid - como por el impacto del consumo y la actividad económica en la generación de residuos, incluyendo también la influencia del turismo y el impacto que la población no local puede tener en dicha generación.

Pese a que el **análisis de regresión y determinantes puede ser de gran utilidad** ya que se centra en identificar relaciones causales entre una variable dependiente (en este caso, la generación de residuos) y un conjunto de variables independientes (sociodemográficas, de actividad económica, y de turismo), este análisis puede estar limitado a la hora de capturar la **complejidad de la relación subyacente entre los datos y la posible interacción entre múltiples variables para concluir sobre la significatividad de las variables analizadas y sus efectos sobre la generación de residuos**.

Es por ello, que en este apartado se propone un ejercicio de análisis multivariante a partir de la técnica de agrupación clúster que recoja el testigo de las principales conclusiones del análisis de determinantes anterior, pero sin limitarse a sus resultados. En este sentido, el **análisis multivariante permite explorar relaciones complejas entre múltiples variables sin necesidad de asumir los supuestos tradicionales de los modelos de regresión** en donde, el análisis clúster, se puede considerar como una técnica fundamental para revelar las relaciones complejas y no lineales entre las variables que puede no ser capturado adecuadamente por los modelos de regresión en el análisis de determinantes.

Para ello, en la primera subsección se realizará una **breve descripción de la técnica** de análisis clúster con las **especificaciones metodológicas** necesarias para, en una segunda

subsección exponer el análisis clúster para la determinación de las zonas homogéneas. Esta identificación **será desarrollada de manera secuencial**, siguiendo la filosofía del análisis en su conjunto, y en la que se **analizará el patrón de determinación de las zonas homogéneas con la inclusión sucesiva de las diferentes variables de interés** (generación de residuos, actividad económica, incidencia del turismo, y características socioeconómicas y demográficas). En la tercera subsección, **se reconciliarán los hallazgos del análisis clúster para proponer la identificación final** de las zonas homogéneas donde realizar la posterior extracción de la muestra.

3.1. Descripción metodológica de la técnica de análisis clúster

Como se ha avanzado anteriormente, el análisis de **clúster es una técnica de agrupamiento que permite identificar patrones naturales o grupos dentro de un conjunto de datos**, sin necesidad de hacer suposiciones previas sobre la estructura de los mismos, a diferencia de los análisis clásicos de regresión.

Esto es especialmente relevante cuando se trabaja con un conjunto diverso de variables, como es este caso, donde se consideran simultáneamente la generación de residuos, variables de tipo socioeconómicas y demográficas, variables relacionadas con la diferente realidad económica de los distritos de la ciudad de Madrid y, sobre la afluencia turística.

En este sentido, **el análisis clúster es fundamental para la posterior extracción de la muestra ya que permite identificar grupos homogéneos de observaciones** para optimizar el diseño muestral y mejorar así la eficiencia en la recopilación de los datos, facilitando una interpretación contextualizada de los resultados sobre la generación de residuos de las actividades económicas.

Es por ello por lo que el **objetivo final del análisis clúster**, como segunda etapa del análisis sistemático y secuencial propuesto, no es sólo identificar las relaciones complejas entre los determinantes de la generación de residuos sino que **utilizar estos patrones para identificar grupos de observaciones**, en este caso de distritos de la ciudad de Madrid, que comparten características similares en relación con estas variables, lo que facilita la segmentación de la población teniendo en cuenta las relaciones subyacentes entre los determinantes de la generación de residuos, y la propia generación.

Es decir, esta segmentación como resultado de la aplicación del análisis clúster va a permitir **identificar un conjunto de zonas que exhibirán comportamientos homogéneos respecto a las variables de interés en el componente intragrupos** (dentro de los clústeres); esto es, respecto a la generación de residuos y su situación socioeconómica y demográfica, su realidad

económica y, la incidencia de la actividad turística en cada distrito, **manteniendo la heterogeneidad entre los grupos identificados.**

Estos grupos o zonas homogéneas representarán la estratificación necesaria de los distritos para proceder, en la siguiente etapa de análisis, con la extracción de la muestra sobre la que se recogerá la información referente a la generación de residuos de las actividades económicas de la Ciudad de Madrid.

Sin embargo, previo a la aplicación de la técnica y el análisis de sus resultados, hay que **resolver dos cuestiones metodológicas fundamentales** circunscritas al caso específico de este estudio: **(i) selección de variables de interés a ser incluidas en el análisis, (ii) selección del tipo de análisis clúster y de su método y medida de agrupación**

Selección de las variables de interés a ser incluidas en el análisis clúster

Como se ha visto en el apartado anterior, **el objeto de este estudio es determinar el grado de generación de residuos de las diferentes actividades económicas** de la ciudad de Madrid dado que esta información no se dispone de manera desagregada para diferenciar qué parte de la generación total se ve afecta por el consumo de los hogares y **qué parte como consecuencia de los usos de la actividad económica** presente en los diferentes distritos de la ciudad, pese a que sí que se conoce este valor a nivel estadístico y para el conjunto de la ciudad (en el entorno del 5% para el año 2023, ver [sección 2.1](#)).

Este hecho hace necesario que el estudio contenga una **visión integral de los factores que pueden influir en la generación de residuos**, tanto de la parte afecta al consumo de los hogares como a la afecta a los resultados de la actividad económica.

En este sentido, aunque se podría realizar una primera *proxy* estadística del valor de la generación de residuos de las actividades económicas aplicando el porcentaje anterior a la variable agregada de carácter administrativo (la que incluye toda la generación de residuos), **esta proxy estaría asumiendo que el porcentaje de residuos afectos a cada distrito es la misma, cuando es sabido que hay variables como la afluencia del sector turístico (mayor infraestructura hotelera según el análisis de determinantes)**, o la presencia de grandes zonas empresariales con menor influencia del consumo de los hogares, que llevaría a errores en la estimación que deberían ser necesarios asumir sin evidencia empírica suficiente como para no rechazarlos.

Así, **la siguiente pregunta** a contestar está relacionada con la **selección de las variables** de interés a ser incluidas en el análisis clúster. Es por ello que en el análisis de determinantes de la etapa anterior, **se decidió consensuadamente** que para determinar esta influencia **se**

debían incluir en el análisis no sólo variables relacionadas con la **diferente estructura económica** de los distritos, sino que también con **las características socioeconómicas y demográficas de sus habitantes para capturar la parte inducida del consumo de los hogares sobre dicha actividad económica, y el mayor o menor grado de desarrollo del sector turístico en cada uno de los mismos.**

En este sentido, **introducir todo el pool de variables en el análisis de clúster**, incluso aquellas que no han resultado significativas en el análisis de determinantes, **puede justificarse por varias razones:**

- **Interacciones entre variables:** Es posible que algunas variables no sean significativas por sí solas, pero podrían interactuar con otras variables de manera significativa, como se vio en el ejemplo de la densidad y el tamaño de la población. De esta manera, al incluir todas las variables en el análisis de clúster, se permite que el algoritmo identifique patrones complejos de interacción que podrían estar presentes entre las variables.
- **Efectos contextuales:** Algunas variables pueden tener efectos contextuales o condicionales que no se capturan en el análisis de determinantes, como por ejemplo, los cambios producidos por la pandemia, cuya significatividad estadística estaba presente en la mayor parte de los modelos. Estos efectos podrían manifestarse únicamente cuando se consideran conjuntamente con otras variables en el análisis de clúster.
- **Reducción de la dimensionalidad de un fenómeno complejo como la generación de residuos:** Aunque algunas variables individuales puedan no ser significativas, podrían contribuir a la identificación de dimensiones latentes o combinaciones de variables que sí son relevantes en el análisis de clúster. Incluir todas las variables en el análisis permite una exploración más exhaustiva de la estructura subyacente de los datos guiada por el sentido económico de las relaciones, y no sólo el estadístico.
- **Evidencia empírica de los modelos estimados:** El ajuste de los modelos estimados aporta evidencias a favor de la selección de **especificaciones funcionales en las que se incluyan todas las variables sin discriminar por su significatividad.** En concreto los criterios de información (AIC y BIC) y el coeficiente de ajuste ($R^2_{Adj.}$) confirmaban un mejor ajuste de las regresiones en el modelo general anidado en el que se incluía todo el pool de variables identificadas como relevantes según el análisis económico previo.

De forma adicional, **los resultados de los estimadores** evidenciaron suficientes relaciones como para concluir que, en mayor o menor medida, **los tres grupos de variables tenían cierta incidencia causal** en la generación de residuos total en los distritos, llevando a la necesidad de mantener los tres grupos de información en las etapas posteriores, aunque no todas las variables de cada grupo se hayan evidenciado como estadísticamente significativas dada la **imposición de la forma funcional lineal que implica el uso de este tipo de modelos econométricos**.

Este supuesto, difícil de asumir dada la complejidad a la hora de explicar el diferente patrón en la generación de residuos, **se ve superado por la aplicación del análisis clúster** que, al realizar un ejercicio de agrupación de carácter multivariante, nos permite **evidenciar patrones no observables por los modelos tradicionales** que expliquen este fenómeno.

En resumen, **se justifica la inclusión de todo el pool de variables identificadas en el análisis de determinantes**, para tratar de capturar a partir del análisis clúster la complejidad de las interacciones entre las variables que podrían influir en la generación de residuos, incluso si algunas de ellas no resultaron significativas de manera individual en el análisis de determinantes. **Esto permitirá una comprensión más completa y detallada de los patrones de agrupamiento de los distritos en función de la generación de residuos**.

Selección del tipo de análisis clúster y de su método y medida de agrupación

La aplicación de la técnica de análisis clúster, igual que el resto de tipos de análisis estadísticos o econométricos, requieren tomar determinadas **decisiones metodológicas previas sobre su especificación**.

En el análisis clúster hay **tres cuestiones** metodológicas a ser resueltas previo a su aplicación: **(i) tipo de agrupación, (ii) método de enlace, y (iii) medida de distancia**. En este ejercicio, se ha determinado realizar un **análisis clúster de tipo jerárquico, con el método de Ward, y teniendo en cuenta las distancias euclídeas al cuadrado**:

- **Clúster jerárquico:** En el análisis de clúster, existen varios enfoques para agrupar observaciones, como **clúster jerárquico y clúster no jerárquico** (k-means). Mientras que el **clúster no jerárquico** requiere especificar de antemano el número de

clústeres deseados, el clúster jerárquico construye una jerarquía de agrupamientos, lo que permite explorar la estructura de los datos a diferentes niveles de granularidad.

Se opta por el clúster jerárquico debido a su capacidad para proporcionar una visión global y estructurada de la relación entre las observaciones, lo que permite identificar patrones de agrupamiento a diferentes niveles de granularidad. Como se verá, **este método permite explorar una amplia gama de posibles estructuras de agrupamiento, desde agrupamientos muy generales hasta agrupamientos más específicos**, lo que proporciona una comprensión completa de la estructura subyacente de los datos.

- **Método de Ward:** Existen **diferentes métodos de enlace**, como el **enlace único completo o promedio**. A diferencia de estos métodos que solo tienen en cuenta la distancia entre dos puntos, **el método de Ward se selecciona debido a su capacidad para minimizar la varianza dentro de cada clúster**, lo que resulta en **clústeres más homogéneos y bien definidos**.

Este método tiende a producir **agrupamientos compactos y de forma regular**, lo que, además **facilita la interpretación de los resultados** y la identificación de patrones claros. En este sentido, el método de Ward es **robusto frente a la presencia de valores atípicos**, lo que lo hace **adecuado para el análisis de conjuntos de datos complejos y multidimensionales**.

- **Distancia euclídea al cuadrado:** En el análisis de clúster, la elección de **la medida de distancia es crucial, ya que afecta directamente a la estructura y la interpretación de los resultados**. Hay varias medidas de distancia disponibles, como la **distancia euclídea**, la **distancia de Manhattan** y la **distancia de Mahalanobis**.

La distancia euclídea al cuadrado se elige como medida de distancia debido a su simplicidad y facilidad de interpretación, así como por ser **ampliamente utilizada en el análisis de clúster debido a su capacidad para capturar la proximidad entre observaciones**. Además, al elevar al cuadrado la distancia euclídea, **se acentúa la diferencia entre las observaciones que están más alejadas entre sí**, lo que puede ayudar a resaltar patrones de agrupamiento más distintivos en los datos.

3.2. Análisis secuencial para la identificación de zonas homogéneas

Con las cuestiones metodológicas resueltas, el objetivo de esta sección es realizar un análisis secuencial aplicando la técnica de análisis clúster especificada y que desemboque en la

identificación de un grupo de zonas (distritos) homogéneas para estratificar la ciudad de Madrid en función de los resultados que se obtengan.

Esta estratificación, es el paso previo y necesario para la correcta extracción de la muestra sobre la que levantar la información referente a la generación de residuos por parte de las actividades económicas.

Para facilitar la comprensión de los resultados finales, la identificación de zonas se realizará por etapas, incluyendo en cada análisis que se realice, un pool adicional de variables identificadas previamente en el siguiente orden:

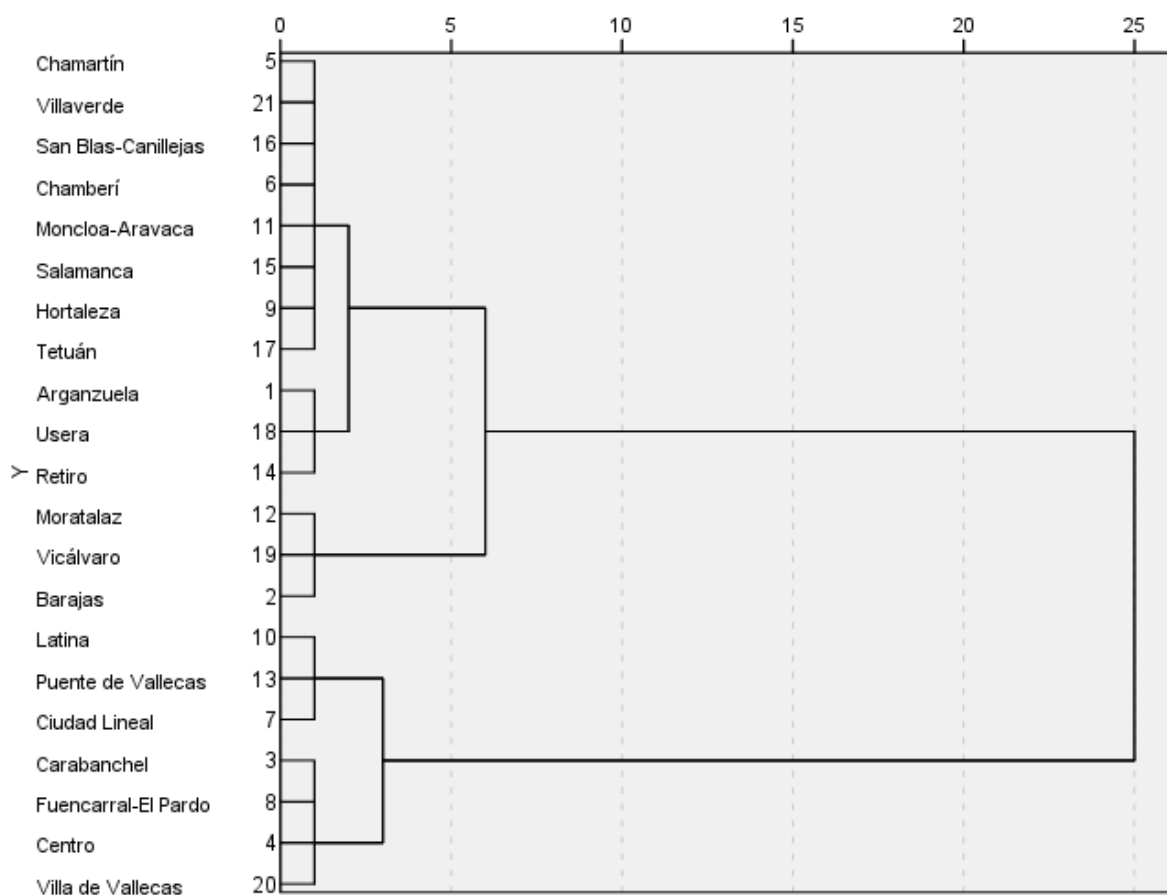
- **Etapla 1.** Análisis de zonas identificadas en función de la generación de residuos
- **Etapla 2.** Etapa 1 + variables referentes a la actividad económica y a la afluencia del sector turístico
- **Etapla 3.** Etapa 2 + variables referentes a la realidad socioeconómica y demográfica de los distritos.

3.2.1. Etapla 1. Zonas homogéneas por generación de residuos

En esta etapa sólo se ha incluido la variable clave del análisis: generación total de residuos, en su versión administrativa.

El resultado del dendrograma es el siguiente:

Figura 10. Identificación zonas homogéneas Etapla 1



Fuente: Elaboración propia

Como se puede observar, **el análisis clúster identifica 5 grupos con un elevado grado de homogeneidad entre ellos** dado que se realizan todas las agrupaciones en el primer nivel de enlace (antes de la primera barrar horizontal que representa la distancia entre los centroides de los grupos/distritos).

Estas zonas son las siguientes:

- **Grupo I.** Chamartín, Villaverde, San Blas-Canillejas, Chamberí, Moncloa-Aravaca, Salamanca, Hortaleza, Tetuán
- **Grupo II.** Arganzuela, Usera, Retiro
- **Grupo III.** Moratalaz, Vicálvaro, Barajas
- **Grupo IV.** Latina, Puente de Vallecas, Ciudad Lineal
- **Grupo V.** Carabanchel, Fuencarral – El Pardo, Centro, Villa de Vallecas

Aunque desde el punto de vista del análisis estadístico el análisis agrupa a los distritos de manera óptima, es imprescindible destacar que al realizarse sólo teniendo en cuenta la generación total de residuos (hogares y actividades económicas) estas agrupaciones pueden

estar capturando de manera sesgada los patrones subyacentes que determinan la generación de residuos.

Este sesgo observacional podría estar determinando agrupaciones difícilmente explicables por el sentido socioeconómico de las agrupaciones, como puede ser que Retiro y Usera se incluyan en el mismo clúster, o Carabanchel y la Villa de Vallecas con Fuencarral-El Pardo y Centro.

Esta cuestión, incide en la necesidad de continuar con el análisis secuencial por etapas para incluir el resto de variables de interés identificadas en los análisis previos.

3.2.2. Etapa 2. Zonas homogéneas por generación de residuos y actividades económicas y del turismo

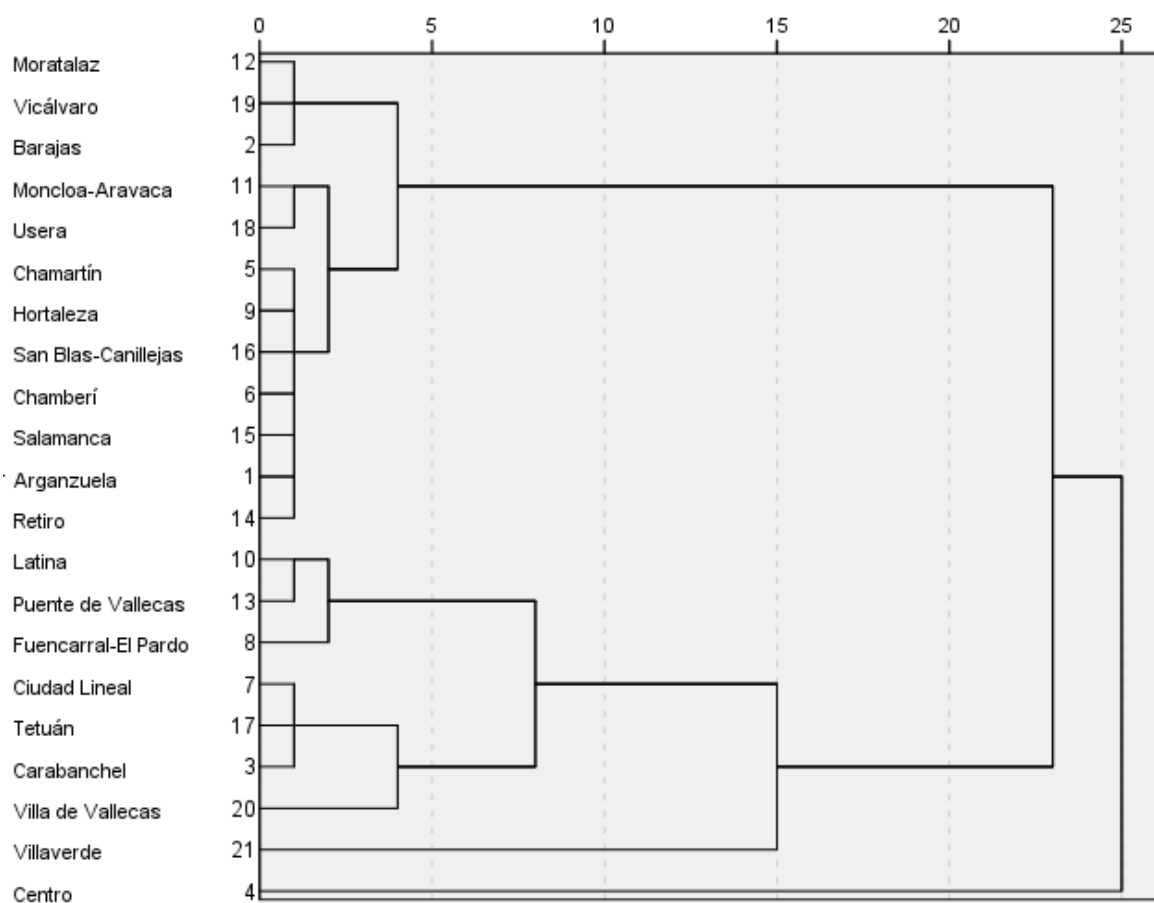
En el análisis de esta etapa, se han incluido las variables de la Etapa 1 (generación total de residuos por distritos), las siguientes variables:

- Variables para capturar la incidencia de la actividad turística en la generación de residuos:
 - % de puntos de interés turístico por distrito
 - % de establecimientos de alojamiento turístico (hoteles, hostales, pensiones)
- Variables para capturar las diferencias entre la actividad económica⁶:
 - % de actividades de construcción
 - % de actividades de industria (agregando extractiva, manufacturera, y suministros)
 - % de actividades de servicios (agregado de actividades profesionales, educación, sanidad, información y telecomunicaciones)
 - % de actividades en los sectores del comercio, la hostelería, y el entretenimiento

El dendrograma obtenido del análisis es el siguiente:

Figura 11. Identificación zonas homogéneas Etapa 2

⁶ En consenso con el Ayuntamiento de Madrid, se excluye la variable referente a la actividad Agricultura debido a que su poca variabilidad por distrito introduce ruido adicional no relevante para la determinación de las zonas homogéneas. Para ello, se ha realizado el análisis de sensibilidad incluyendo y sin incluir la variable agricultura y justificando su exclusión en la poca representación de los porcentajes de actividad económica en cada distrito, y dada la poca relevancia del sector agricultura en la ciudad de Madrid (8 locales en la ciudad que se dedican a la actividad de este sector) y en la generación de las fracciones de residuos que se analiza.



Fuente: Elaboración propia

Como se puede observar, **la introducción de las variables de actividades económicas y zonas de interés turístico ha introducido cierta heterogeneidad** entre los grupos, lo que, a su vez, ha llevado a la identificación de **zonas con un único distrito (Villaverde o Centro)**, dada la dificultad de obtener cierta **homogeneidad intradistritos** para establecer la agrupación si tenemos en cuenta la variabilidad del nuevo conjunto de información.

Es importante destacar que, pese a la heterogeneidad incluida por la inclusión de las nuevas variables, se **conserva parte de la homogeneidad establecida en la identificación de grupos de la Etapa 1** como por ejemplo el grupo identificado para Moratalaz, Vicálvaro y Barajas.

En definitiva, la técnica ha llevado a la identificación de siete zonas con diferente grado de homogeneidad (agrupaciones en diferentes niveles), que han sido los siguientes:

- **Grupo I.** Moratalaz, Vicálvaro, Barajas (Grupo III anterior)
- **Grupo II.** Moncloa-Aravaca, Usara

- **Grupo III.** Chamartín, Hortaleza, San Blas-Canillejas, Chamberí, Salamanca, Arganzuela, Retiro
- **Grupo IV.** Latina, Puente de Vallecas, Fuencarral – El Pardo (2º nivel de agregación)
- **Grupo V.** Ciudad Lineal, Tetuán, Carabanchel, Villa de Vallecas (3º nivel de agregación)
- **Grupo VI.** Villaverde
- **Grupo VII.** Centro

3.2.3. Etapa 3. Zonas homogéneas con la inclusión de todo el pool de variables

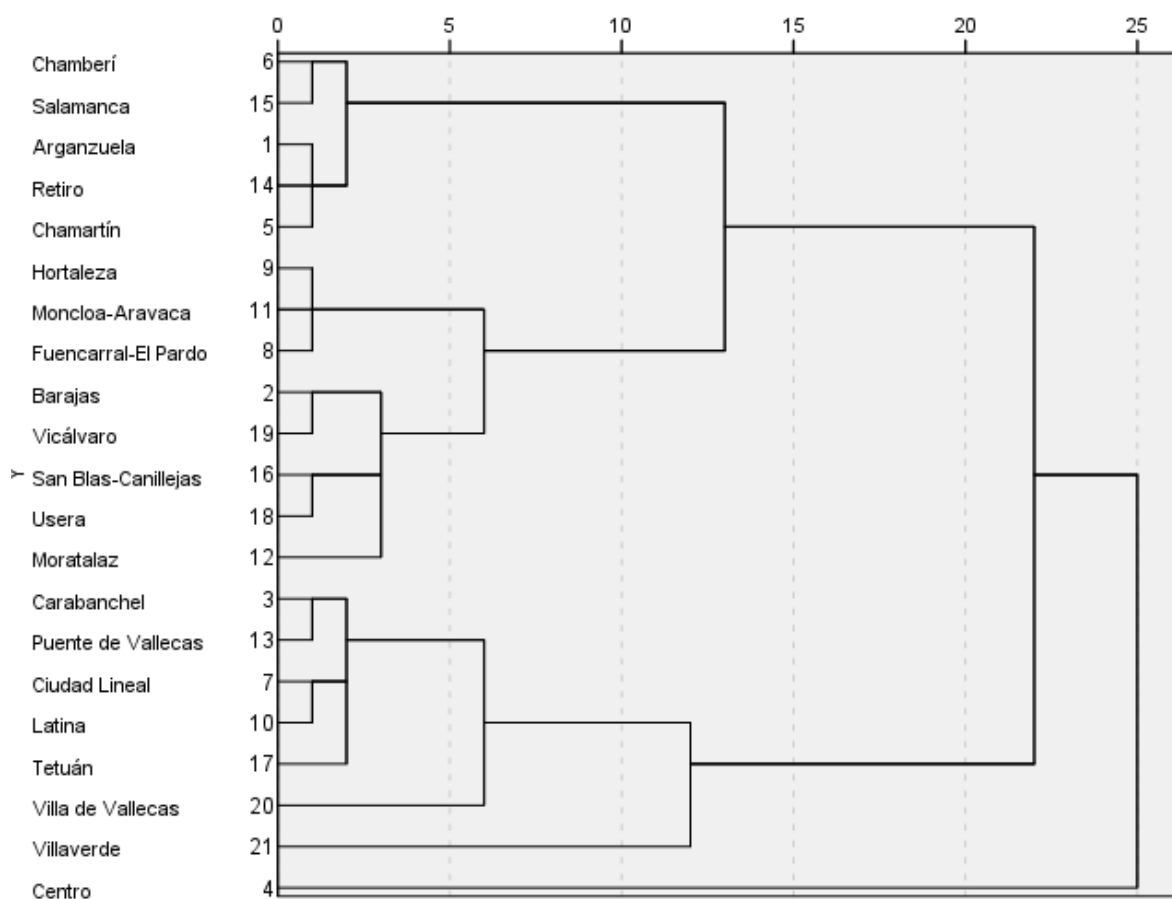
En esta etapa, las variables que se han incluido son las identificadas en la Etapa 2, más aquellas referidas a la caracterización socioeconómica y demográfica de los distritos de la ciudad de Madrid. En concreto, las variables adicionales incluidas en esta etapa son las siguientes:

- Densidad de población
- Índice de envejecimiento
- Renta disponible media
- Precio medio de la vivienda

Tal y como se puede observar en la Figura 12, la variabilidad introducida por la realidad socioeconómica y demográfica de los distritos ha permitido consolidar las zonas homogéneas identificadas a partir de dos puntos de interés:

- Reducción del número final de clúster, incluidos los dos con un único distrito (6)
- Retención de un mayor número de clústeres en el primer nivel de agrupación (mayor homogeneidad intragrupos)

Figura 12. Identificación zonas homogéneas Etapa 3



Fuente: Elaboración propia

El resultado de las zonas homogéneas identificadas es el siguiente:

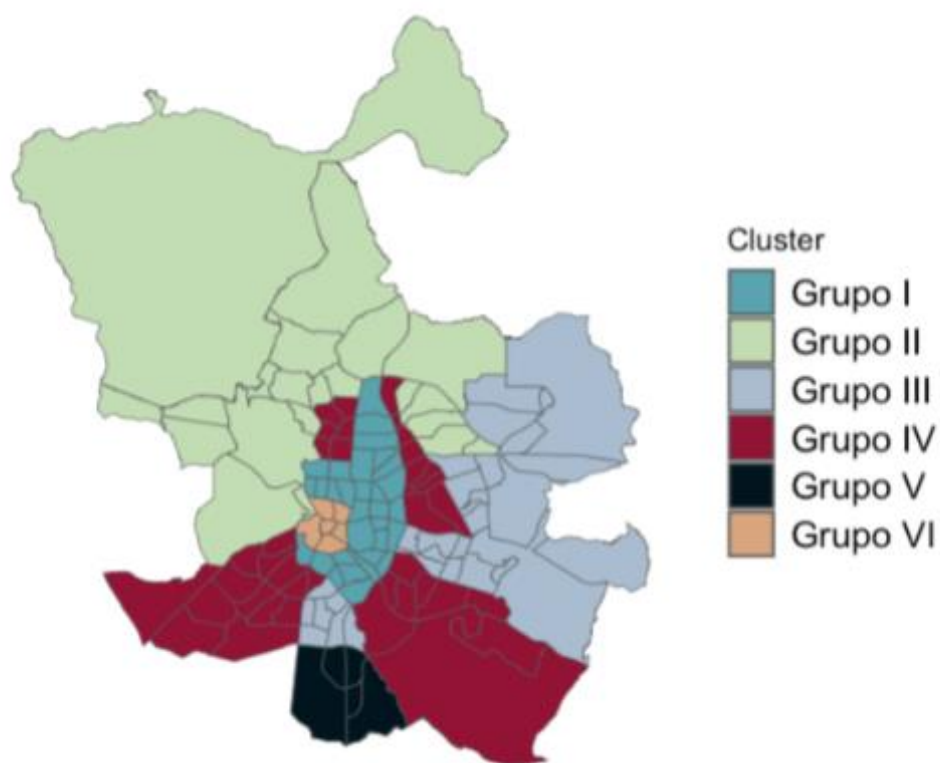
- **Grupo I.** Chamberí, Salamanca, Arganzuela, Retiro, Chamartín
- **Grupo II.** Hortaleza, Moncloa, Fuencarral
- **Grupo III.** Barajas, Vicálvaro, Canillejas, Usera, Moratalaz
- **Grupo IV.** Carabanchel, Puente de Vallecas, Ciudad Lineal, Latina, Tetuán, Villa de Vallecas (3era agregación)
- **Grupo V.** Villaverde
- **Grupo VI.** Centro

Como se puede observar, las zonas obtenidas en esta ocasión tienen un **mayor sentido económico teniendo en cuenta la diferente realidad social y económica de los distritos de la ciudad de Madrid**, y en particular capturando el distinto grado de homogeneidad intragrupos y heterogeneidad entre grupos tanto de la actividad turística como de la actividad económica, socioeconómica y demográfica.

El mejor **ejemplo** de ello es la **unicidad de los grupos referidos a Villaverde** (98,8% de la industria extractiva, el 12% de la construcción, el 93,4% de los suministros, o el 97% de las comunicaciones) **o Centro** (44% de las zonas turísticas, 66% de los establecimientos turísticos y el 14% de las actividades de comercio, hostelería y entretenimiento) **o la agrupación** del Grupo I (Chamberí, Salamanca, Arganzuela, Retiro y Chamartín) **delimitando geográficamente la zona centro de la ciudad**, al igual que el Grupo II (Hortaleza, Moncloa, Fuencarral).

Para identificar visualmente la **pertinencia de las zonas homogéneas identificadas** a partir de la clusterización, la Figura 13 muestra el **mapa coroplético** para los diferentes grupos.

Figura 13. Mapa coroplético de zonas homogéneas identificadas



Fuente: Elaboración propia

4. Muestreo aleatorio estratificado por zonas homogéneas

El punto de partida para el trabajo de muestreo es la información contenida en dos archivos facilitados por el Ayuntamiento de Madrid: Listado_AE y Marco_Trúa, que contienen las bases de datos del Impuesto sobre actividades económicas y la Tasa de Residuos urbanos respectivamente. Junto a estos, de modo complementario, se analizaron los archivos, también proporcionados por el Ayuntamiento, con el registro de los contenedores que tienen los locales empresariales.

Comenzaremos con el análisis descriptivo de estos dos archivos, los cruces de información entre ellos y las principales conclusiones para la realización del muestreo.

4.1. Análisis descriptivo de los ficheros de trabajo iniciales.

4.1.1. Análisis descriptivo de Listado_AE y proceso de depuración

Tabla 16. Estructura de “Listado_AE”

NOMBRE	DÍGITOS	DESCRIPCIÓN
Referencia Catastral	18	Referencia Catastral
Local	9	Código del Local
Código_Actividad	6	Código de la actividad
Epígrafe	6	Código del epígrafe (CNAE 4 Dígitos) ⁽¹⁾
Division_Epígrafe	2	Código de la división del epígrafe (CNAE 2 Dígitos) ⁽²⁾
Seccion_Epígrafe	1	Código de la sección del epígrafe (CNAE -Letra- 1 dígito) ⁽²⁾
DEC_Actividad	TXT	Descripción del Código de actividad
DEC_Epígrafe	TXT	Descripción del Epígrafe
DEC_Division	TXT	Descripción de la División
DEC_Seccion	TXT	Descripción de la Sección
Rotulo	TXT	Rótulo del establecimiento
Agrupacion	TXT	Agrupación del Local

(1) Aunque tiene 6 dígitos, parece acercarse al código CNAE a 4 dígitos, aunque tiene una estructura variable (en algunos casos es CNAE a 4 dígitos a los que se añaden dos dígitos, que comienzan en 01 y avanzan sucesivamente con mayor o menor desarrollo según la División_Epígrafe en el que estemos; en otros casos no se añaden estos dos códigos detrás, sino que se incluyen dos ceros en la posición cuatro y cinco).

(2) Coincidencia total con CNAE

Contenido de “Listado_AE”

En total, los registros contenidos en el fichero son 121.648. Analizado por Referencia Catastral hay considerables duplicidades, es decir varias actividades asociadas a una misma Referencia Catastral. La siguiente tabla clasifica las Referencias catastrales en función del número de actividades (registros del fichero) al que están asociadas.

Tabla 17. Número de actividades por referencia catastral

N.º de Actividades	Listado AE	
	Absolutos	%
1	79.368	83,98%
2	10.715	11,34%
3	2.579	2,73%
4	702	0,74%
5-9	882	0,93%
10-14	143	0,15%
15-19	60	0,06%
20-49	45	0,05%
50-99	6	0,01%
100-199	2	0,00%
200 y +	1	0,00%
TOTAL	94.503	100,00%

Por tanto, contamos con un total de 94.503 Referencias Catastrales que agrupan 121.648 actividades diferentes (registros).

El análisis de las Referencias Catastrales con varias actividades asociadas muestra que, básicamente hay dos tipos de repeticiones: Referencias Catastrales que reúnen en un único establecimiento varias actividades y Referencias Catastrales que reúnen distintos establecimientos, cada uno de ellos con su actividad.

En las siguientes páginas mostramos varios ejemplos de estos dos tipos de repeticiones.

Ejemplo 1. Referencia Catastral con ocho actividades. Supermercado del que se desagrega cada una de las categorías de producto comercializado.

R. Catastral	Local	Codigo_Actividad	Epigrafe	Division_Epigrafe	Seccion_Epigrafe	DEC_Actividad	DEC_Epigrafe
0441702VK4084A0181	270542898	204441	471101	47	G	SUPERMERCADO	COMERCIO MENOR NO ESPECIALIZADO, CON PREDOMINIO EN ALIMENTOS
0441702VK4084A0181	270542898	204441	472102	47	G	SUPERMERCADO	COMERCIO MENOR DE FRUTAS Y HORTALIZAS
0441702VK4084A0181	270542898	204441	472202	47	G	SUPERMERCADO	COMERCIO MENOR DE CHARCUTERIA
0441702VK4084A0181	270542898	204441	472204	47	G	SUPERMERCADO	COMERCIO MENOR DE CARNICERIA-SALCHICHERIA
0441702VK4084A0181	270542898	204441	472206	47	G	SUPERMERCADO	COMERCIO MENOR DE AVES, HUEVOS Y CAZA
0441702VK4084A0181	270542898	204441	472302	47	G	SUPERMERCADO	COMERCIO MENOR DE PESCADOS Y MARISCOS
0441702VK4084A0181	270542898	204441	472401	47	G	SUPERMERCADO	COMERCIO MENOR DE PAN Y BOLLERIA CON OBRADOR
0441702VK4084A0181	270542898	204441	472407	47	G	SUPERMERCADO	ESTABLECIMIENTOS DE PLATOS PREPARADOS CON OBRADOR

Ejemplo 2. Referencia catastral con ocho actividades. Diferentes actividades sin relación, excepto en el caso de una peluquería que se desagrega en Peluquería y Centro de estética.

R. Catastral	Local	Codigo_Actividad	Epigrafe	Division_Epigrafe	Seccion_Epigrafe	DEC_Actividad	DEC_Epigrafe
0152803VK4705C0001	10001775	170657	472102	47	G	COMERCIO MENOR FRUTAS Y HORTALIZAS	COMERCIO AL POR MENOR DE FRUTAS Y HORTALIZAS SIN OBRADOR
0152803VK4705C0001	270537070	195309	563005	56	I	BAR SIN COCINA	BAR SIN COCINA
0152803VK4705C0001	280011174	202847	855002	85	P	ENSEÑANZA NO REGLADA	ENSEÑANZA NO REGLADA (DEPORTIVA RECREATIVA, CULTURAL)
0152803VK4705C0001	280011172	220570	960201	96	S	PELUQUERÍA	SERVICIO DE PELUQUERIA
0152803VK4705C0001	280011172	220570	960203	96	S	PELUQUERÍA	CENTRO DE ESTETICA
0152803VK4705C0001	280014264	337926	730001	73	M	PUBLICIDAD, RELACIONES PUBLICAS	PUBLICIDAD, RELACIONES PUBLICAS Y ESTUDIOS DE MERCADO
0152803VK4705C0001	300002699	339352	591001	59	J	PRODUCTORA	ACTIVIDADES CINEMATOGRAFICAS, DE VIDEO Y DE TELEVISION
0152803VK4705C0001	270508192	8154	476101	47	G	LIBRERÍA	COMERCIO AL POR MENOR DE LIBROS

En los ejemplos se ve claramente como el código de local establece diferencias entre actividades que pertenecen al mismo establecimiento (idéntico código de local) y actividades que, aunque se realizan en un espacio con una única Referencia Catastral, pertenecen a empresas diferentes (distintos códigos de local).

Eliminación de locales repetidos

Por tanto, partiendo del código de local se eliminaron todos los locales repetidos. Tras este proceso de depuración el número total de registros en el fichero Listado_AE es de 106.302.

4.1.2. Análisis descriptivo de MARCO_TRUA y proceso de depuración.

Tabla 18. Estructura de “marco TRUA”

NOMBRE	DÍGITOS	DESCRIPCIÓN
Referencia Catastral	18	Referencia Catastral
Código del Distrito	2	Código Distrito municipal
Código del Barrio	3	Código Barrio
Naturaleza del Inmueble	TXT	Rústica/Urbana/BICES
Uso Catastral	TXT	12 usos catastrales
Dirección Total	TXT	Dirección postal completa
Dirección	TXT	Tipo de vía / Nombre de Vía / (Número, km.)
Código de la vía	8	Código (tamaño variable 1-8)
NDP	8	Código (tamaño variable 5-8)
Código de Local	6	Código (tamaño variable 3-6)
Tipo de Vía	TXT	Tipo de Vía
Nombre de la Vía	TXT	Nombre de la vía
Bloque	TXT	Bloque
Escalera	TXT	Escalera
Planta	TXT	Planta
Puerta	TXT	Puerta
Código Postal	5	Código postal
Coordenada1	7	Coordenada 1
Coordenada2	7	Coordenada 2
Código Control RC	TXT	Dos letras (¿?)
Superficie Construida	7	Metros cuadrados? (hasta 7 dígitos)
Superficie Gráfica	8	Metros cuadrados? (hasta 8 dígitos)
NIF	9	NIF del titular
Personalidad		Física/Jurídica
Titular_catastral		Denominación del titular

Contenido de “MARCO_TRUA”

El fichero contiene un total de 148.755 registros, correspondiendo cada uno de ellos a una Referencia Catastral diferente.

Todos los registros cuentan con información suficiente sobre su ubicación. Más allá de la información relacionada con la dirección del local, en el fichero hay información relacionada con el tipo de inmueble y con características de su propiedad. Las siguientes tablas recogen la información que contienen estas variables.

Tabla 19. Descriptivos de “MARCO_TRUA”

NATURALEZA INMUEBLE	
BICES	31
Rústica	25
Urbana	148.699
TOTAL	148.755

USO CATASTRAL	
AEROPUERTOS Y PUERTOS COMERC.	30
COMERCIAL	96.509
CULTURAL	2.248
DEPORTIVO	1.223
EDIFICIO SINGULAR	371
ESPECTACULOS	110
INDUSTRIAL	10.862
OCIO Y HOSTELERIA	5.425
OFICINAS	29.956
PRESAS SALTOS AGUA EMBALSES	1
RELIGIOSO	721
SANIDAD Y BENEFICENCIA	1.299
TOTAL	148.755

SUPERFICIE CONSTRUIDA	
Con información	148.719
Sin información	36
TOTAL	148.755

PERSONALIDAD	
FÍSICA	80.341
JURÍDICA	68.401
Sin información	13
TOTAL	148.755

Sólo en el caso de que la titularidad sea de una persona jurídica el fichero incluye el NIF y la denominación del titular. Es decir, el 54% de los locales carecen de ninguna identificación del titular.

4.1.3. Análisis descriptivo de los listados de contenedores

Además de los ficheros tratados hasta aquí, el Ayuntamiento proporcionó varios ficheros que contenían información sobre los contenedores para distintos residuos proporcionados por el Ayuntamiento a algunos establecimientos. Concretamente, los ficheros que contenían esta información se denominaban: MADL4_DEPENDENCIAS_2403_L1.xlsx, MADL4_DEPENDENCIAS_2403_L2.xlsx y MADL4_DEPENDENCIAS_2403_L3.xlsx.

En estos ficheros para cada contenedor entregado (registro) aparece la siguiente información:

- Específica del contenedor: Fracción, Tipología y Modelo.
- Información de ubicación del local al que pertenece el contenedor: Distrito, Barrio, Dirección (calle y número), coordenadas (x,y).

La única variable común en los listados de contenedores y Marco_TRUA es la dirección (incompleta) que dificulta enormemente poder contar con esta información en el estudio. Trabajando con un ejemplo concreto se evidencia esta dificultad: seleccionado una calle con toda su numeración en ambas bases de datos podemos ejemplificar con claridad los problemas que se plantean.

En la siguiente tabla mostramos un ejemplo con una calle completa.

Tabla 20. Ejemplo con una calle completa

	ESTABLECIMIENTOS (TRUA)	CONTENEDORES (MADL4DEPENDENCIAS)
CALLE MONTERA	3	0
CALLE MONTERA 1	1	0
CALLE MONTERA 3	1	1
CALLE MONTERA 4	3	0
CALLE MONTERA 5	2	2
CALLE MONTERA 6	1	0
CALLE MONTERA 8	0	1
CALLE MONTERA 9	1	0
CALLE MONTERA 10	1	1
CALLE MONTERA 11	2	0
CALLE MONTERA 13	4	1
CALLE MONTERA 14	3	0
CALLE MONTERA 15	6	0
CALLE MONTERA 16	1	0
CALLE MONTERA 19	1	0
CALLE MONTERA 20	9	0
CALLE MONTERA 21	2	0

	ESTABLECIMIENTOS (TRUA)	CONTENEDORES (MADL4DEPENDENCIAS)
CALLE MONTERA 22	1	0
CALLE MONTERA 24	102	0
CALLE MONTERA 25	0	3
CALLE MONTERA 26	6	0
CALLE MONTERA 28	1	0
CALLE MONTERA 30	4	0
CALLE MONTERA 32	42	0
CALLE MONTERA 33	14	0
CALLE MONTERA 34	1	4
CALLE MONTERA 36	1	1
CALLE MONTERA 37	1	0
CALLE MONTERA 38	3	0
CALLE MONTERA 39	3	0
CALLE MONTERA 40	1	0
CALLE MONTERA 41	5	0
CALLE MONTERA 42	4	2
CALLE MONTERA 43	4	0
CALLE MONTERA 45	1	0
CALLE MONTERA 47	3	40
TOTAL	238	56

En verde los establecimientos con contenedores que no aparecen en “Marco_TRUA” y en naranja los contenedores de imposible asignación porque en la dirección hay varios establecimientos.

Por tanto, de los 56 contenedores asignados a algún negocio en la Calle Montero:

- 45 contenedores con graves dificultades para asignar: hay varios establecimientos y no sabemos a cuál/es asignar los distintos contenedores.
- 4 contenedores imposibles de asignar: en la dirección no hay ningún establecimiento en el listado TRUA.
- 7 contenedores para los que hay un único establecimiento, los podríamos asignar siempre que supongamos que todos los contenedores pertenecen a ese establecimiento.

Es decir, en este ejemplo, como mucho podríamos asignar un 13% de los contenedores.

Sólo podríamos asignar contenedores en aquellos establecimientos que son únicos en un número de una vía. En el registro de establecimientos de “Marco_TRUA” están en esta situación el 31% de los registros. Este sería el máximo teórico porque si tomamos como referencia el

ejemplo de la calle Montera mostrado anteriormente: de los 14 teóricamente asignables (un único establecimiento) sólo en 4 casos había contenedores para asignar. Si esta proporción se extendiera a todo el municipio finalmente se asignarían contenedores a un 9% de los registros.

4.1.4. Creación del marco muestral: enlace de ficheros.

Tras la depuración del fichero “Listado_AE”, eliminando todos aquellos locales “repetidos” (número de local repetido) se procedió a la asignación de toda la información de ubicación contenido en el fichero “Marco_TRUA” a cada uno de los registros.

De los 106.302 registros del “Listado_AE” no repetidos, 76.835 tienen un número de Registro Catastral coincidente con los registros de ubicación (“Marco_TRUA”).

La estructura del fichero de establecimientos resultante, atendiendo a su actividad y a su ubicación (agrupación de distritos en 6 Grupos homogéneos) es el siguiente:

Tabla 21: Número de establecimientos por cada grupo homogéneo

Actividad	Grupos Homogéneos (distritos)						TOTAL
	G1	G2	G3	G4	G5	G6	
A) AGRICULTURA, GANADERÍA, SILVICULTURA Y PESCA	6	1	4	4		2	17
B) INDUSTRIAS EXTRACTIVAS	1			1	1		3
C) INDUSTRIA MANUFACTURERA	330	209	257	1.054	128	133	2.111
D) SUMINISTRO DE ENERGÍA ELÉCTRICA, GAS, VAPOR Y AIRE ACONDICIONADO	35	33	32	31	7	15	153
E) SUMINISTRO DE AGUA, ACTIVIDADES DE SANEAMIENTO, GESTIÓN DE RESIDUOS Y DESCONTAMINACIÓN	10	4	9	15	1	1	40
F) CONSTRUCCIÓN	450	220	246	706	71	64	1.757
G) COMERCIO AL POR MAYOR Y AL POR MENOR	7.152	2.604	2.965	8.927	967	3.069	25.684
H) TRANSPORTE Y ALMACENAMIENTO	348	169	236	553	88	106	1.500
I) HOSTELERÍA	3.290	1.536	1.498	3.729	405	2.452	12.910
J) INFORMACIÓN Y COMUNICACIONES	513	214	230	546	40	167	1.710
K) ACTIVIDADES FINANCIERAS Y DE SEGUROS	822	259	220	627	54	168	2.150
L) ACTIVIDADES INMOBILIARIAS	491	153	128	383	49	115	1.319
M) ACTIVIDADES PROFESIONALES, CIENTÍFICAS Y TÉCNICA	1.842	488	312	1.049	68	469	4.228
N) ACTIVIDADES ADMINISTRATIVAS Y SERVICIOS AUXILIARES	1.797	534	417	1.222	68	396	4.434
O) ADMINISTRACIÓN PÚBLICA	155	80	33	102	15	82	467
P) EDUCACIÓN	915	481	377	868	72	247	2.960
Q) ACTIVIDADES SANITARIAS Y DE SERVICIOS SOCIALES	1.498	585	477	1.206	115	180	4.061
R) ACTIVIDADES ARTÍSTICAS, RECREATIVAS Y DE ENTRETENIMIENTO	487	205	238	528	52	258	1.768
S) OTROS SERVICIOS	2.317	955	1.039	2.871	299	684	8.165

Actividad	Grupos Homogéneos (distritos)						TOTAL
	G1	G2	G3	G4	G5	G6	
U) ACTIVIDADES DE ORGANIZACIONES Y ORGANISMOS EXTRATERRITORIALES	56	7	2	12		8	85
Z) SIN ACTIVIDAD	377	201	132	465	23	115	1.313
TOTAL	22.892	8.938	8.852	24.899	2.523	8.731	76.835

Es necesario eliminar los 1.313 registros que tienen codificada como actividad el epígrafe Z (inexistente en la CNAE) y que se define en el fichero como “sin actividad”. Eliminando estos registros, tomaríamos como universo del estudio los 75.522 establecimientos.

4.2. Proceso de definición de la muestra

El análisis de la estructura y contenido del universo del que partimos plantea varios problemas para la distribución de la muestra. Recordemos que se pretende alcanzar una muestra de 1.200 establecimientos y que el objetivo fundamental es estimar los residuos de distintos tipos que se generan en los seis grupos de distritos por cada actividad.

Los principales problemas son:

- Hay un número excesivamente elevado de celdillas (combinaciones actividad/grupo homogéneo): en total son 120, aunque en 5 de ellas no hay ningún establecimiento.
- Hay cinco actividades con un número muy pequeño de establecimientos. Estas cinco actividades concentran, en total, 298 establecimientos (0,4% del total). Concretamente se trata de:
 - A) Agricultura, ganadería, silvicultura y pesca (17 establecimientos)
 - B) Industrias extractivas (3 establecimientos)
 - D) Suministro de energía eléctrica, gas, vapor y aire acondicionado (153 establecimientos)
 - E) Suministro de agua, actividades de saneamiento, gestión de residuos y descontaminación (40 establecimientos)
 - U) Actividades de organizaciones y organismos extraterritoriales (85 establecimientos)

La división por actividades está definida en función del producto o servicio principal del establecimiento. Esto implica una diversidad mucho más amplia que la que teóricamente supone la generación de los residuos que vamos a analizar (recordemos que están fuera todos los residuos con un tratamiento especial, que únicamente analizaremos: papel-cartón, vidrio, envases, orgánica y resto). Desde esta perspectiva, revisando todas las actividades,

proponemos agrupar todas las actividades que se desarrollan en oficinas, independientemente de la actividad concreta que realicen. Concretamente, proponemos la agrupación de los siguientes epígrafes:

Actividades que se desarrollan en oficinas

- J) Información y comunicaciones
- K) Actividades financieras y de seguros
- L) Actividades inmobiliarias
- M) Actividades profesionales, científicas y técnicas
- N) Actividades administrativas y servicios auxiliares
- O) Administración pública
- U) Actividades de organizaciones y organismos extraterritoriales

Se plantea una duda con la agrupación M) Actividades profesionales, científicas y técnicas, por la inclusión de laboratorios y otro tipo de establecimientos donde se desarrollan actividades científicas con características específicas. Detengámonos en la estructura detallada de esta agrupación.

Tabla 22: Estructura detallada de la agrupación M

SECCIÓN M		Grupo H.						TOTAL
División		G1	G2	G3	G4	G5	G6	
69	ACTIVIDADES JURÍDICAS Y DE CONTABILIDAD	741	124	109	422	39	153	1.588
70	ACTIVIDADES DE LAS SEDES CENTRALES	277	81	47	139	6	57	607
71	SERVICIOS TÉCNICOS DE ARQUITECTURA E INGENIERÍA	266	87	36	136	9	66	600
72	INVESTIGACIÓN Y DESARROLLO	20	22	7	14		3	66
73	PUBLICIDAD Y ESTUDIOS DE MERCADO	204	50	23	85	2	84	448
74	OTRAS ACTIVIDADES PROFESIONALES, CIENTÍFICAS Y TÉCNICAS	265	68	49	166	6	97	651
75	ACTIVIDADES VETERINARIAS	69	56	41	87	6	9	268
TOTAL		1.842	488	312	1.049	68	469	4.228

Las actividades marcadas en rojo (72. Investigación y desarrollo y 75. Actividades veterinarias) serían las que deberíamos excluir de la agrupación generada para “actividades desarrolladas en oficinas”. Teniendo en cuenta que representan un número reducido de establecimientos y, sobre todo, por las similitudes en cuanto a los residuos que generan con las actividades sanitarias decidimos integrarlas en este grupo.

La agrupación Oficinas solventa uno de los dos problemas planteados: el excesivo número de actividades (celdillas), sin embargo, apenas incide en solucionar el problema generado por las actividades con un número muy reducido de establecimiento (sólo la sección U. Actividades de organizaciones y organismos extraterritoriales se integra en la agrupación de oficinas).

Teniendo en cuenta que del sector primario sólo hay 17 establecimientos y que el resto de las actividades con pocos establecimientos podrían agruparse en el sector Industrial creemos que la mejor solución es generar un Grupo que integrara ambos sectores.

Aplicando estas dos soluciones agregativas (oficinas, por un lado y sector primario e industria, por otro), el universo quedaría del siguiente modo:

Tabla 23: Distribución del universo de partida

	Grupos Homogéneos (distritos)						TOTAL
	G1 (5)	G2 (3)	G3 (5)	G4 (6)	G5 (1)	G6 (1)	
A-E) AGRICULTURA, INDUSTRIA Y SUMINISTROS (ENERGÍA/AGUA)	382	247	302	1.105	137	151	2.324
F) CONSTRUCCIÓN	450	220	246	706	71	64	1.757
G) COMERCIO AL POR MAYOR Y AL POR MENOR	7.152	2.604	2.965	8.927	967	3.069	25.684
H) TRANSPORTE Y ALMACENAMIENTO	348	169	236	553	88	106	1.500
I) HOSTELERÍA	3.290	1.536	1.498	3.729	405	2.452	12.910
J-O,U) ACTIVIDADES QUE SE DESARROLLAN EN UNA OFICINA (excl. 72 y 75)	5.587	1.657	1.294	3.840	288	1.393	14.059
P) EDUCACIÓN	915	481	377	868	72	247	2.960
Q) ACTIVIDADES SANITARIAS Y DE SERVICIOS SOCIALES (inc. 72 y 75)	1.587	663	525	1.307	121	192	4.395
R) ACTIVIDADES ARTÍSTICAS, RECREATIVAS Y DE ENTRETENIMIENTO	487	205	238	528	52	258	1.768
S) OTROS SERVICIOS	2.317	955	1.039	2.871	299	684	8.165
TOTAL	22.515	8.737	8.720	24.434	2.500	8.616	75.522

(Las sombreadas son actividades agrupadas)

4.2.1 Alternativa 1. Distribución proporcional de la muestra

La primera opción de distribución es **respetar las proporciones** que representan las actividades y grupos homogéneos en el universo. La distribución que obtenemos es la siguiente:

Tabla 24: Distribución proporcional de la muestra

Alternativa 1 Distribución Proporcional	Grupos Homogéneos (distritos)						TOTAL
	G1 (5)	G2 (3)	G3 (5)	G4 (6)	G5 (1)	G6 (1)	
A-E) AGRICULTURA, INDUSTRIA Y SUMINISTROS (ENERGÍA/AGUA)	6	4	5	18	2	2	37
F) CONSTRUCCIÓN	7	3	4	11	1	1	27
G) COMERCIO AL POR MAYOR Y AL POR MENOR	113	41	47	142	15	49	407

H) TRANSPORTE Y ALMACENAMIENTO	6	3	4	9	1	2	25
I) HOSTELERÍA	52	24	24	59	6	39	204
J-O,U) ACTIVIDADES QUE SE DESARROLLAN EN UNA OFICINA (excl. 72 y 75)	89	26	21	61	5	22	224
P) EDUCACIÓN	15	8	6	14	1	4	48
Q) ACTIVIDADES SANITARIAS Y DE SERVICIOS SOCIALES (inc. 72 y 75)	25	11	8	21	2	3	70
R) ACTIVIDADES ARTÍSTICAS, RECREATIVAS Y DE ENTRETENIMIENTO	8	3	4	8	1	4	28
S) OTROS SERVICIOS	37	15	16	46	5	11	130
TOTAL	358	138	139	389	39	137	1.200

Esta distribución nos ofrece un error muestral para los datos a nivel ciudad de $\pm 2,9$ puntos⁷.

Tabla 25: Errores asociados a la distribución proporcional

Errores Alternativa 1	Grupos Homogéneos (distritos)						TOTAL
	G1 (5)	G2 (3)	G3 (5)	G4 (6)	G5 (1)	G6 (1)	
A-E) AGRICULTURA, INDUSTRIA Y SUMINISTROS (ENERGÍA/AGUA)	40,6%	49,7%	44,4%	23,4%	70,5%	70,5%	16,3%
F) CONSTRUCCIÓN	37,5%	57,5%	49,7%	29,9%	100,0%	100,0%	19,1%
G) COMERCIO AL POR MAYOR Y AL POR MENOR	9,3%	15,5%	14,5%	8,3%	25,6%	14,2%	4,9%
H) TRANSPORTE Y ALMACENAMIENTO	40,5%	57,4%	49,7%	33,1%	100,0%	70,4%	19,8%
I) HOSTELERÍA	13,8%	20,3%	20,3%	12,9%	40,6%	15,9%	6,9%
J-O,U) ACTIVIDADES QUE SE DESARROLLAN EN UNA OFICINA (exc. 72 y 75)	10,5%	19,5%	21,7%	12,7%	44,4%	21,2%	6,6%
P) EDUCACIÓN	25,6%	35,1%	40,6%	26,5%	100,0%	49,7%	14,3%
Q) ACTIVIDADES SANITARIAS Y DE SERVICIOS SOCIALES (inc. 72 y 75)	19,8%	29,9%	35,1%	21,7%	70,4%	57,4%	11,9%
R) ACTIVIDADES ARTÍSTICAS, RECREATIVAS Y DE ENTRETENIMIENTO	35,1%	57,5%	49,7%	35,1%	100,0%	49,7%	18,8%
S) OTROS SERVICIOS	16,3%	25,6%	24,8%	14,6%	44,4%	29,9%	8,7%
TOTAL	5,2%	8,4%	8,4%	5,0%	15,9%	8,5%	2,9%

Esta distribución sería la más adecuada si quisiéramos obtener información para el conjunto de la ciudad. Sin embargo, la información para cada grupo homogéneo es muy irregular, sobre todo en el caso del Grupo 5 (en el que está en solitario el distrito de Villaverde) donde tendríamos que realizar estimaciones basadas en 39 encuestas (error $\pm 15,9$). Además, los errores se disparan en cada celdilla, llegando en algunas de ellas a alcanzar el error máximo (100).

⁷ Calculados partiendo de los criterios del muestreo aleatorio simple, en el caso de mayor incertidumbre posible p y q = 50. Margen de confianza 95,5%.

4.2.2. Alternativa 2. Distribución homogénea de la muestra

La segunda opción sería realizar una distribución homogénea de la muestra por Grupo para, posteriormente dentro de cada Grupo, distribuir proporcionalmente las entrevistas a realizar.

Esta alternativa se materializa en la siguiente distribución:

Tabla 26: Distribución homogénea por Grupo

Alternativa 2	Grupos Homogéneos (distritos)						TOTAL
Distribución homogénea por Grupo H.	G1 (5)	G2 (3)	G3 (5)	G4 (6)	G5 (1)	G6 (1)	
A-E) AGRICULTURA, INDUSTRIA Y SUMINISTROS (ENERGÍA/AGUA)	3	6	7	9	11	4	40
F) CONSTRUCCIÓN	4	5	6	6	6	1	28
G) COMERCIO AL POR MAYOR Y AL POR MENOR	64	59	68	73	77	72	413
H) TRANSPORTE Y ALMACENAMIENTO	3	4	5	5	7	2	26
I) HOSTELERÍA	29	35	34	31	32	57	218
J-O, U) ACTIVIDADES QUE SE DESARROLLAN EN UNA OFICINA (excl. 72 y 75)	50	38	30	31	23	32	204
P) EDUCACIÓN	8	11	9	7	6	6	47
Q) ACTIVIDADES SANITARIAS Y DE SERVICIOS SOCIALES (inc. 72 y 75)	14	15	12	11	10	4	66
R) ACTIVIDADES ARTÍSTICAS, RECREATIVAS Y DE ENTRETENIMIENTO	4	5	5	4	4	6	28
S) OTROS SERVICIOS	21	22	24	23	24	16	130
TOTAL	200	200	200	200	200	200	1.200

Con esta distribución todos los datos a nivel Grupo Homogéneo tendrían errores cercanos a ± 7 puntos. Por otro lado, como se muestra en la siguiente tabla, los errores en cada una de las celdillas, aunque mejoran, continúan siendo muy elevados.

Tabla 26: Errores asociados a la distribución homogénea

Errores Alternativa 2	Grupos Homogéneos (distritos)					
Distribución homogénea por Grupo H.	G1 (5)	G2 (3)	G3 (5)	G4 (6)	G5 (1)	G6 (1)
A-E) AGRICULTURA, INDUSTRIA Y SUMINISTROS (ENERGÍA/AGUA)	57,7%	40,8%	37,8%	33,3%	30,1%	50,0%
F) CONSTRUCCIÓN	50,0%	44,7%	40,8%	40,8%	40,8%	100,0%
G) COMERCIO AL POR MAYOR Y AL POR MENOR	12,5%	13,0%	12,1%	11,7%	11,2%	11,7%
H) TRANSPORTE Y ALMACENAMIENTO	57,7%	50,0%	44,7%	44,7%	37,8%	70,7%
I) HOSTELERÍA	18,6%	16,9%	17,1%	17,9%	17,6%	13,2%
J-O, U) ACTIVIDADES QUE SE DESARROLLAN EN UNA OFICINA (excl. 72 y 75)	14,1%	16,2%	18,2%	17,9%	20,8%	17,6%
P) EDUCACIÓN	35,3%	30,1%	33,3%	37,8%	40,8%	40,8%

Q) ACTIVIDADES SANITARIAS Y DE SERVICIOS SOCIALES (inc. 72 y 75)	26,7%	25,8%	28,8%	30,1%	31,6%	50,0%
R) ACTIVIDADES ARTÍSTICAS, RECREATIVAS Y DE ENTRETENIMIENTO	50,0%	44,7%	44,7%	50,0%	50,0%	40,8%
S) OTROS SERVICIOS	21,8%	21,3%	20,4%	20,8%	20,3%	25,0%
TOTAL	7,0%	7,0%	7,0%	7,0%	6,8%	7,0%

4.2.3. Alternativa 3. Distribución estableciendo un mínimo

Teniendo en cuenta que el objetivo final es realizar una estimación de la producción de residuos para cada agrupación de actividad y cada grupo homogéneo, es decir, a nivel celdilla, la mejor opción sería asegurar un número de entrevistas mínimo en cada una de ellas.

Para ello, el procedimiento se compone de dos pasos. Un primer paso en el que se asigna un número idéntico a cada celdilla. Asignando 10 entrevistas a cada celda distribuiríamos el 50% de la muestra (600).

Tabla 27: Distribución estableciendo un mínimo (paso 1)

Alternativa 3. Paso 1	Grupos Homogéneos (distritos)						TOTAL
Mínimo homogéneo	G1 (5)	G2 (3)	G3 (5)	G4 (6)	G5 (1)	G6 (1)	
A-E) AGRICULTURA, INDUSTRIA Y SUMINISTROS (ENERGÍA/AGUA)	10	10	10	10	10	10	60
F) CONSTRUCCIÓN	10	10	10	10	10	10	60
G) COMERCIO AL POR MAYOR Y AL POR MENOR	10	10	10	10	10	10	60
H) TRANSPORTE Y ALMACENAMIENTO	10	10	10	10	10	10	60
I) HOSTELERÍA	10	10	10	10	10	10	60
J-O,U) ACTIVIDADES QUE SE DESARROLLAN EN UNA OFICINA (exc. 72 y 75)	10	10	10	10	10	10	60
P) EDUCACIÓN	10	10	10	10	10	10	60
Q) ACTIVIDADES SANITARIAS Y DE SERVICIOS SOCIALES (inc. 72 y 75)	10	10	10	10	10	10	60
R) ACTIVIDADES ARTÍSTICAS, RECREATIVAS Y DE ENTRETENIMIENTO	10	10	10	10	10	10	60
S) OTROS SERVICIOS	10	10	10	10	10	10	60
TOTAL	100	100	100	100	100	100	600

El segundo paso sería distribuir las 600 encuestas restantes de modo homogéneo según los Grupos y, dentro de ellos, de modo proporcional según la Actividad.

Tabla 28: Distribución estableciendo un mínimo (paso 2)

Alternativa 3. Paso 2 Resto: homogéneo por grupo y proporcional por actividad	Grupos Homogéneos (distritos)						TOTAL
	G1 (5)	G2 (3)	G3 (5)	G4 (6)	G5 (1)	G6 (1)	
A-E) AGRICULTURA, INDUSTRIA Y SUMINISTROS (ENERGÍA/AGUA)	2	3	3	5	5	2	20
F) CONSTRUCCIÓN	2	3	3	3	3	1	15
G) COMERCIO AL POR MAYOR Y AL POR MENOR	32	30	34	36	39	36	207
H) TRANSPORTE Y ALMACENAMIENTO	2	2	3	2	4	1	14
I) HOSTELERÍA	14	17	17	15	16	28	107
J-O,U) ACTIVIDADES QUE SE DESARROLLAN EN UNA OFICINA (exc. 72 y 75)	25	19	15	16	11	16	102
P) EDUCACIÓN	4	5	4	4	3	3	23
Q) ACTIVIDADES SANITARIAS Y DE SERVICIOS SOCIALES (inc. 72 y 75)	7	8	6	5	5	2	33
R) ACTIVIDADES ARTÍSTICAS, RECREATIVAS Y DE ENTRETENIMIENTO	2	2	3	2	2	3	14
S) OTROS SERVICIOS	10	11	12	12	12	8	65
TOTAL	100	100	100	100	100	100	600

La distribución final sería la suma de los resultados de los dos pasos. La siguiente tabla muestra el resultado.

Tabla 29: Distribución final alternativa elegida

Alternativa 3. Distribución final Mínimo por celdilla	Grupos Homogéneos (distritos)						TOTAL
	G1 (5)	G2 (3)	G3 (5)	G4 (6)	G5 (1)	G6 (1)	
A-E) AGRICULTURA, INDUSTRIA Y SUMINISTROS (ENERGÍA/AGUA)	12	13	13	15	15	12	80
F) CONSTRUCCIÓN	12	13	13	13	13	11	75
G) COMERCIO AL POR MAYOR Y AL POR MENOR	42	40	44	46	49	46	267
H) TRANSPORTE Y ALMACENAMIENTO	12	12	13	12	14	11	74
I) HOSTELERÍA	24	27	27	25	26	38	167
J-O, U) ACTIVIDADES QUE SE DESARROLLAN EN UNA OFICINA (excl.. 72 y 75)	35	29	25	26	21	26	162
P) EDUCACIÓN	14	15	14	14	13	13	83
Q) ACTIVIDADES SANITARIAS Y DE SERVICIOS SOCIALES (inc. 72 y 75)	17	18	16	15	15	12	93

R) ACTIVIDADES ARTÍSTICAS, RECREATIVAS Y DE ENTRETENIMIENTO	12	12	13	12	12	13	74
S) OTROS SERVICIOS	20	21	22	22	22	18	125
TOTAL	200	200	200	200	200	200	1.200

Esta distribución hace que contemos desde 11 entrevistas para hacer la estimación de residuos que generan las empresas de Construcción y Transporte en G6, hasta más de 40 para los Comercios de todos los Grupos.

Si calculamos el error estándar en cada estrato:

$$SE_i = \frac{SD_i}{\sqrt{N_{itt}}}$$

Como muestra la siguiente tabla, los errores oscilan entre $\pm 15,1$ y $\pm 7,1$, mucho más bajos que cualquiera de las otras alternativas planteadas previamente.

Tabla 30: Errores asociados a la alternativa elegida de selección de muestra

Errores Alternativa 3 Mínimo por celdilla	Grupos Homogéneos (distritos)					
	G1 (5)	G2 (3)	G3 (5)	G4 (6)	G5 (1)	G6 (1)
A-E) AGRICULTURA, INDUSTRIA Y SUMINISTROS (ENERGÍA/AGUA)	14,4%	13,9%	13,9%	12,9%	12,9%	14,4%
F) CONSTRUCCIÓN	14,4%	13,9%	13,9%	13,9%	13,9%	15,1%
G) COMERCIO AL POR MAYOR Y AL POR MENOR	7,7%	7,9%	7,5%	7,4%	7,1%	7,4%
H) TRANSPORTE Y ALMACENAMIENTO	14,4%	14,4%	13,9%	14,4%	13,4%	15,1%
I) HOSTELERÍA	10,2%	9,6%	9,6%	10,0%	9,8%	8,1%
J-O, U) ACTIVIDADES QUE SE DESARROLLAN EN UNA OFICINA (excl. 72 y 75)	8,5%	9,3%	10,0%	9,8%	10,9%	9,8%
P) EDUCACIÓN	13,4%	12,9%	13,4%	13,4%	13,9%	13,9%
Q) ACTIVIDADES SANITARIAS Y DE SERVICIOS SOCIALES (inc. 72 y 75)	12,1%	11,8%	12,5%	12,9%	12,9%	14,4%
R) ACTIVIDADES ARTÍSTICAS, RECREATIVAS Y DE ENTRETENIMIENTO	14,4%	14,4%	13,9%	14,4%	14,4%	13,9%
S) OTROS SERVICIOS	11,2%	10,9%	10,7%	10,7%	10,7%	11,8%

4.2.4. Conclusiones sobre la distribución de la muestra.

El objetivo final del estudio, realizar una estimación sobre la generación de residuos por actividad para las distintas zonas en las que se ha dividido la ciudad (Grupos Homogéneos), hace que la distribución que mejor se adapte sea la última alternativa, la que establece un número mínimo de encuestas para cada cruce Actividad por Grupo homogéneo.

Por tanto, el siguiente cuadro muestra la distribución que utilizaremos.

Tabla 31: Alternativa elegida de selección de muestra

Alternativa 3. Distribución final Mínimo por celdilla	Grupos Homogéneos (distritos)						TOTAL
	G1 (5)	G2 (3)	G3 (5)	G4 (6)	G5 (1)	G6 (1)	
A-E) AGRICULTURA, INDUSTRIA Y SUMINISTROS (ENERGÍA/AGUA)	12	13	13	15	15	12	80
F) CONSTRUCCIÓN	12	13	13	13	13	11	75
G) COMERCIO AL POR MAYOR Y AL POR MENOR	42	40	44	46	49	46	267
H) TRANSPORTE Y ALMACENAMIENTO	12	12	13	12	14	11	74
I) HOSTELERÍA	24	27	27	25	26	38	167
J-O, U) ACTIVIDADES QUE SE DESARROLLAN EN UNA OFICINA (excl. 72 y 75)	35	29	25	26	21	26	162
P) EDUCACIÓN	14	15	14	14	13	13	83
Q) ACTIVIDADES SANITARIAS Y DE SERVICIOS SOCIALES (inc. 72 y 75)	17	18	16	15	15	12	93
R) ACTIVIDADES ARTÍSTICAS, RECREATIVAS Y DE ENTRETENIMIENTO	12	12	13	12	12	13	74
S) OTROS SERVICIOS	20	21	22	22	22	18	125
TOTAL	200	200	200	200	200	200	1.200

4.3. Proceso de extracción de la muestra

La **primera opción para la extracción de la muestra barajada es la extracción de los registros partiendo del “Listado_AE”** cualificado con la ubicación geográfica y el resto de la información **contenida en el “Marco_TRUA”**. Sin embargo, la extracción directamente de este fichero plantea diversos problemas:

- 1) Inexistencia de teléfonos asociados. Teniendo en cuenta que las entrevistas se deben realizar telefónicamente la falta de esta información requiere un proceso de búsqueda complejo.
- 2) Sólo se cuenta con la identificación de los propietarios de los locales cuando estos son personas jurídicas. Es decir, el 54% de los locales carecen de ninguna identificación del titular. Para la extracción de la muestra tendríamos que dejar fuera a este 54% de los locales de los que son titulares personas físicas.
- 3) Aunque nos limitáramos a extraer las personas jurídicas y, de estas consiguiéramos sus teléfonos de contacto, esto implicaría contactar con el propietario del establecimiento

que, en una proporción muy elevada, no se corresponderá con el titular de la empresa que desarrolla la actividad en él.

Por tanto, **realizar la extracción directa de este fichero implicaría no sólo dejar a la mayoría de los locales fuera de la muestra, sino que, además, generaría graves dificultades para la obtención de la información necesaria.**

Como solución, hemos optado por **realizar una petición a una empresa dedicada al suministro de bases de datos de empresas.** El universo disponible de negocios en la ciudad de Madrid en este caso asciende a 333.780 registros, conformando un universo rico de datos que más que triplica el número de registros que contiene la base de datos de la Tasa de Residuos⁸. La petición realizada es la siguiente:

- 1) 1.200 empresas y autónomos de la ciudad de Madrid con la distribución establecida en la muestra teórica. En cada una de las celdillas es necesario que se realice una selección aleatoria entre todas las empresas y autónomos asignados en la celdilla. Este fichero con 1.200 empresas y autónomos reproducirá fielmente la estructura de la muestra teórica y será la que conforme la **muestra titular**.

Tabla 32: Distribución de la muestra titular

DISTRIBUCIÓN DE LA MUESTRA Muestra titular	Grupos Homogéneos (distritos)						TOTAL
	G1 (5)	G2 (3)	G3 (5)	G4 (6)	G5 (1)	G6 (1)	
A-E) AGRICULTURA, INDUSTRIA Y SUMINISTROS (ENERGÍA/AGUA)	12	13	13	15	15	12	80
F) CONSTRUCCIÓN	12	13	13	13	13	11	75
G) COMERCIO AL POR MAYOR Y AL POR MENOR	42	40	44	46	49	46	267
H) TRANSPORTE Y ALMACENAMIENTO	12	12	13	12	14	11	74
I) HOSTELERÍA	24	27	27	25	26	38	167
J-O,U) ACTIVIDADES QUE SE DESARROLLAN EN UNA OFICINA (exc. 72 y 75)	35	29	25	26	21	26	162
P) EDUCACIÓN	14	15	14	14	13	13	83
Q) ACTIVIDADES SANITARIAS Y DE SERVICIOS SOCIALES (inc. 72 y 75)	17	18	16	15	15	12	93
R) ACTIVIDADES ARTÍSTICAS, RECREATIVAS Y DE ENTRETENIMIENTO	12	12	13	12	12	13	74
S) OTROS SERVICIOS	20	21	22	22	22	18	125
TOTAL	200	200	200	200	200	200	1.200

⁸ Esto es así porque la tasa está diseñada para cobrarse por la actividad principal del inmueble, pudiendo un inmueble albergar más de una actividad.

- 2) Una **muestra sustituta** compuesta por nueve registros por cada registro titular. Es decir, extracción de 10.800 registros con la siguiente distribución:

Tabla 33: Distribución de la muestra sustituta

Muestra sustituta	Grupos Homogéneos (distritos)						TOTAL
	G1 (5)	G2 (3)	G3 (5)	G4 (6)	G5 (1)	G6 (1)	
A-E) AGRICULTURA, INDUSTRIA Y SUMINISTROS (ENERGÍA/AGUA)	108	117	117	135	135	108	720
F) CONSTRUCCIÓN	108	117	117	117	117	99	675
G) COMERCIO AL POR MAYOR Y AL POR MENOR	378	360	396	414	441	414	2.403
H) TRANSPORTE Y ALMACENAMIENTO	108	108	117	108	126	99	666
I) HOSTELERÍA	216	243	243	225	234	342	1.503
J-O, U) ACTIVIDADES QUE SE DESARROLLAN EN UNA OFICINA (excl. 72 y 75)	315	261	225	234	189	234	1.458
P) EDUCACIÓN	126	135	126	126	117	117	747
Q) ACTIVIDADES SANITARIAS Y DE SERVICIOS SOCIALES (inc. 72 y 75)	153	162	144	135	135	108	837
R) ACTIVIDADES ARTÍSTICAS, RECREATIVAS Y DE ENTRETENIMIENTO	108	108	117	108	108	117	666
S) OTROS SERVICIOS	180	189	198	198	198	162	1.125
TOTAL	1.800	1.800	1.800	1.800	1.800	1.800	10.800

Para generar este segundo fichero de registros sustitutos el procedimiento de extracción será idéntico al de la muestra titular: aleatorio entre todas las empresas y autónomos existentes en cada cruce de actividad*grupo homogéneo, excluyendo aquellos que formen parte de la muestra titular.

4.4. Nueva distribución tras la eliminación de los Grandes Productores.

Tras la decisión del Ayuntamiento de Madrid de dejar fuera del universo del estudio a lo que se cataloga como Grandes Productores y de los que proporcionó una enumeración en tres ficheros diferentes⁹, fue necesaria su eliminación del marco muestral de partida y, por tanto, la repetición del procedimiento de distribución.

A pesar de que suponemos que el efecto de la eliminación de estos Grandes Productores será mínimo (por un lado, su número es reducido y por otro, la alternativa muestral por la que se ha

⁹ Lote 1_2023 11 03 GRANDES PRODUCTORES_Lote 1 1feb24.xlsx
Lote 2_Grandes_Productores_Coordenadas.xlsx
L3 Calculo tasa Grandes productores.xlsx

optado – número mínimo por celdilla – minimizará el impacto) decidimos ajustar el universo y volver a realizar la distribución. Las siguientes tablas recogen los resultados.

Tabla 34. Distribución del universo de partida (Sin Grandes Productores)

	Grupos Homogéneos (distritos)						TOTAL
	G1 (5)	G2 (3)	G3 (5)	G4 (6)	G5 (1)	G6 (1)	
A-E) AGRICULTURA, INDUSTRIA Y SUMINISTROS (ENERGÍA/AGUA)	382	245	301	885	136	151	2.100
F) CONSTRUCCIÓN	450	220	246	701	71	64	1.752
G) COMERCIO AL POR MAYOR Y AL POR MENOR	7.152	2.558	2.886	8.568	967	3.069	25.200
H) TRANSPORTE Y ALMACENAMIENTO	348	169	233	546	88	106	1.490
I) HOSTELERÍA	3.289	1.514	1.471	3.696	405	2.447	12.822
J-O, U) ACTIVIDADES QUE SE DESARROLLAN EN UNA OFICINA (excl. 72 y 75)	5.587	1.653	1.288	3.783	288	1.392	13.991
P) EDUCACIÓN	915	481	377	867	72	247	2.959
Q) ACTIVIDADES SANITARIAS Y DE SERVICIOS SOCIALES (inc. 72 y 75)	1.585	661	525	1.302	121	192	4.386
R) ACTIVIDADES ARTÍSTICAS, RECREATIVAS Y DE ENTRETENIMIENTO	487	205	233	520	52	257	1.754
S) OTROS SERVICIOS	2.317	945	1.030	2.853	299	683	8.127
TOTAL	22.512	8.651	8.590	23.721	2.499	8.608	74.581

En total, tras la eliminación de los grandes productores, hay 941 registros menos que en el universo. Los establecimientos de Mercamadrid, uno de los grandes productores, es el que tiene un mayor impacto, de hecho el 76% de los establecimientos eliminados del universo inicial se concentran en el Grupo Homogéneo 4, en el que está el distrito donde se ubica Mercamadrid.

La distribución final, después de asignar un mínimo de 10 encuestas a cada celdilla y asignar el resto de las pendientes hasta las 200 asignadas a cada Grupo Homogéneo de modo proporcional por actividad, es la siguiente:

Tabla 35. Distribución final de la muestra (sin Grandes Productores)

Alternativa 3bis. Distribución final Mínimo por celdilla sin GP	Grupos Homogéneos (distritos)						TOTAL
	G1 (5)	G2 (3)	G3 (5)	G4 (6)	G5 (1)	G6 (1)	
A-E) AGRICULTURA, INDUSTRIA Y SUMINISTROS (ENERGÍA/AGUA)	12	13	13	14	15	12	79
F) CONSTRUCCIÓN	12	13	13	13	13	11	75

Alternativa 3bis. Distribución final	Grupos Homogéneos (distritos)						TOTAL
Mínimo por celdilla sin GP	G1 (5)	G2 (3)	G3 (5)	G4 (6)	G5 (1)	G6 (1)	
G) COMERCIO AL POR MAYOR Y AL POR MENOR	42	39	44	46	49	46	266
H) TRANSPORTE Y ALMACENAMIENTO	12	12	13	12	14	11	74
I) HOSTELERÍA	24	27	27	26	26	38	168
J-O, U) ACTIVIDADES QUE SE DESARROLLAN EN UNA OFICINA (excl. 72 y 75)	35	29	25	26	21	26	162
P) EDUCACIÓN	14	16	14	14	13	13	84
Q) ACTIVIDADES SANITARIAS Y DE SERVICIOS SOCIALES (inc. 72 y 75)	17	18	16	15	15	12	93
R) ACTIVIDADES ARTÍSTICAS, RECREATIVAS Y DE ENTRETENIMIENTO	12	12	13	12	12	13	74
S) OTROS SERVICIOS	20	21	22	22	22	18	125
TOTAL	200	200	200	200	200	200	1.200

Por lo que respecta a los errores asociados, la eliminación de los Grandes Productores prácticamente no ha producido variaciones en la muestra anterior y, por tanto, también se mantienen casi inalterables los errores (Tabla 31).

Tabla 36: Errores asociados a la distribución sin Grandes Productores

Errores Alternativa 3bis	Grupos Homogéneos (distritos)					
Mínimo por celdilla sin GP	G1 (5)	G2 (3)	G3 (5)	G4 (6)	G5 (1)	G6 (1)
A-E) AGRICULTURA, INDUSTRIA Y SUMINISTROS (ENERGÍA/AGUA)	14,4%	13,9%	13,9%	13,4%	12,9%	14,4%
F) CONSTRUCCIÓN	14,4%	13,9%	13,9%	13,9%	13,9%	15,1%
G) COMERCIO AL POR MAYOR Y AL POR MENOR	7,7%	8,0%	7,5%	7,4%	7,1%	7,4%
H) TRANSPORTE Y ALMACENAMIENTO	14,4%	14,4%	13,9%	14,4%	13,4%	15,1%
I) HOSTELERÍA	10,2%	9,6%	9,6%	9,8%	9,8%	8,1%
J-O, U) ACTIVIDADES QUE SE DESARROLLAN EN UNA OFICINA (excl. 72 y 75)	8,5%	9,3%	10,0%	9,8%	10,9%	9,8%
P) EDUCACIÓN	13,4%	12,5%	13,4%	13,4%	13,9%	13,9%
Q) ACTIVIDADES SANITARIAS Y DE SERVICIOS SOCIALES (inc. 72 y 75)	12,1%	11,8%	12,5%	12,9%	12,9%	14,4%
R) ACTIVIDADES ARTÍSTICAS, RECREATIVAS Y DE ENTRETENIMIENTO	14,4%	14,4%	13,9%	14,4%	14,4%	13,9%
S) OTROS SERVICIOS	11,2%	10,9%	10,7%	10,7%	10,7%	11,8%

